

University of Groningen

Computation of characteristics of value-of-time distributions and their standard errors

Meijer, Erik

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2004

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Meijer, E. (2004). *Computation of characteristics of value-of-time distributions and their standard errors*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Computation of characteristics of value-of-time distributions and their standard errors

Erik Meijer*

April 29, 2004

SOM-theme F Interactions between consumers and firms

Abstract

It is discussed how from some estimated bivariate distributions characteristics of the distribution of the ratio of the two random variables and standard errors thereof can be computed.

I would like to thank Jan Rouwendal, Ton Steerneman, and Tom Wansbeek for stimulating discussions and helpful comments.

* Department of Econometrics, University of Groningen, P.O. Box 800, 9700 AV Groningen, THE NETHERLANDS. E-mail: e.meijer@eco.rug.nl.

1 Introduction

Meijer and Rouwendal (2000, 2004) estimated the parameters of the multivariate distribution of the random coefficients in a mixed logit model, based on various distributional assumptions. From these estimated multivariate distributions, the corresponding estimated distributions of the value of time (VOT) have been derived, where the VOT is the ratio of the second and first coefficients. Here, it is discussed how estimates of several characteristics of the distribution of the VOT and standard errors corresponding to these estimates can be computed.

Section 2 discusses the various model specifications and the resulting distributions and densities of the VOT. Section 3 discusses the computation of various characteristics of these VOT distributions. Section 4 explains how standard errors can be computed and sections 5, 6, and 7 discuss some more complicated subproblems of the computation of the standard errors. In section 8, the related problem of constructing confidence bands for density functions and confidence regions for probability mass functions is briefly discussed. Section 9 briefly discusses the empirical results. Finally, section 10 discusses the existence of the characteristics of one of the submodels.

2 The models and corresponding VOT distributions and densities

In the models, the parameters of the multivariate distribution of a vector β of random coefficients have been estimated by (simulated) maximum likelihood. Let $F_\beta(\beta; \theta)$ be the distribution function of the vector β , which is a known parametric function of β and a parameter vector θ . The vector β consists of (random) coefficients corresponding to the travel cost, travel time, number of interchanges, and comfort level of a certain journey by train, which are the weights of these variables in a linear conditional indirect utility function as part of a (mixed) logit model.

The models that have been estimated are based on the following distributional assumptions:

- I** Standard logit, no random coefficients;
- II** Normally distributed random coefficients;
- III** Lognormally distributed random coefficients;

IV Gamma distributed random coefficients;

V Latent class approach: random coefficients follow a discrete distribution with $J = 9$ mass points.

For the models II–IV, four variants have been estimated, based on a cross-classification of whether the cost coefficient is nonrandom or random and whether the random coefficients are independently distributed or not:

- (a) Cost coefficient nonrandom, random coefficients independent;
- (b) Cost coefficient nonrandom, random coefficients dependent;
- (c) Cost coefficient random, random coefficients independent;
- (d) Cost coefficient random, random coefficients dependent.

For variants (b) and (d), the dependence is introduced by letting the random coefficients be a multivariate normally distributed vector, univariately transformed to the desired marginal distributions. The rationale for the various models and variants can be found in the original papers. Further details can also be found there.

The value of time is a derived random variable, denoted by r ,

$$r = \beta_2/\beta_1.$$

We are interested in the sensitivity of the distribution of r to assumptions about the distribution F_β of β . Clearly, the distribution of r depends only on the bivariate marginal distribution of $(\beta_1, \beta_2)'$, or, equivalently, on the subset of θ that defines this bivariate distribution. Therefore, with a slight abuse of notation and without loss of generality, we let $F_\beta(\beta; \theta)$ be the bivariate distribution function of $\beta = (\beta_1, \beta_2)'$, where θ is the vector consisting of the parameters of this bivariate distribution.

In model I, both coefficients are nonrandom and the value of time is a constant, $r = \beta_2/\beta_1$.

In model II, the coefficients are bivariate normally distributed with means μ_1 and μ_2 , respectively, variances σ_1^2 and σ_2^2 , covariance σ_{12} , and resulting correlation $\rho = \sigma_{12}/(\sigma_1\sigma_2)$. In variants (a) and (b), σ_1^2 is zero and ρ is undefined. The cost coefficient β_1 is therefore equal to μ_1 with probability one, and hence we will treat it as a constant and refer to this parameter as β_1 instead of μ_1 . In variant (c), σ_1^2 is allowed to be nonzero, but $\rho = 0$. The result is that in variants (a) and (b), r is normally distributed

with mean (μ_2/β_1) and variance σ_2^2/β_1^2 . In variants (c) and (d), the density of r is

$$g(r) = \frac{\sqrt{1-\rho^2}}{\pi\sigma_1\sigma_2a^2(r)} \exp\left[-\frac{c}{2(1-\rho^2)}\right] + \frac{b(r)d(r)}{\sqrt{2\pi}\sigma_1\sigma_2a^3(r)} \left[\Phi\left[\frac{b(r)}{a(r)\sqrt{1-\rho^2}}\right] - \Phi\left[-\frac{b(r)}{a(r)\sqrt{1-\rho^2}}\right] \right], \quad (1)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and

$$\begin{aligned} a(r) &= \left(\frac{r^2}{\sigma_2^2} - \frac{2\rho r}{\sigma_1\sigma_2} + \frac{1}{\sigma_1^2} \right)^{1/2}, \\ b(r) &= \frac{\mu_2 r}{\sigma_2^2} - \frac{\rho(\mu_2 + \mu_1 r)}{\sigma_1\sigma_2} + \frac{\mu_1}{\sigma_1^2}, \\ c &= \frac{\mu_2^2}{\sigma_2^2} - \frac{2\rho\mu_1\mu_2}{\sigma_1\sigma_2} + \frac{\mu_1^2}{\sigma_1^2}, \\ d(r) &= \exp\left[\frac{b^2(r) - ca^2(r)}{2(1-\rho^2)a^2(r)}\right]. \end{aligned}$$

Of course, in variant (c), these formulas can be simplified somewhat because $\rho = 0$. The distribution function is

$$G(r) = \int_{-\infty}^r g(u) du. \quad (2)$$

In practice, this function can be approximated by simulation: a large number B , say, of drawings of $(\beta_1, \beta_2)'$ from the estimated bivariate normal distribution are generated, giving $(\beta_{1b}, \beta_{2b})'$, $b = 1, \dots, B$ and $r_b = \beta_{2b}/\beta_{1b}$. An estimate of $G(r)$ is then

$$\hat{G}_{\text{sim}}(r) = \frac{1}{B} \sum_{b=1}^B I(r_b \leq r).$$

This is a step function with step size $1/B$ at r_b , $b = 1, \dots, B$. By choosing B large enough, this function approximates (2) arbitrarily close. Alternatively, $G(r)$ can be approximated by some (other) form of numerical integration.

The model as estimated is a reparameterized version, where not the covariance matrix Σ of the random coefficients is estimated, but its Cholesky root L , so $\Sigma = LL'$ and L is a lower triangular matrix. This leads to an easier estimation algorithm and ensures that the estimated covariance matrix is indeed a covariance matrix, i.e., positive semidefinite. For variants (a) and (b), where the monetary coefficient is nonrandom, the

corresponding row and column of L are fixed to zero, so that the corresponding row and column of Σ are zero as well. For variants (a) and (c), where all coefficients are independent, the off-diagonal elements of L are fixed to zero, so that L is a diagonal matrix and thus Σ is diagonal as well. These details are necessary to mention explicitly because they play a role in the computation of the standard errors in section 4.

In model III, the coefficients are (-1) times lognormally distributed random variables, such that the natural logarithms of these lognormally distributed random variables are bivariate normally distributed with means μ_1 and μ_2 , respectively, variances σ_1^2 and σ_2^2 , and covariance σ_{12} . The same reparameterizations as in model II are used in this model. In all four variants, r is lognormally distributed. Its parameters are $(\mu_2 - \mu_1, \sigma_2^2)$ in variants (a) and (b), $(\mu_2 - \mu_1, \sigma_1^2 + \sigma_2^2)$ in variant (c), and $(\mu_2 - \mu_1, \sigma_1^2 + \sigma_2^2 - 2\sigma_{12})$ in variant (d). Note that in this case $\beta_1 = -\exp(\mu_1)$ in variants (a) and (b).

In models IV(a) and (b), the cost coefficient is a constant, β_1 , and the time coefficient is Gamma distributed with shape parameter α_2 and scale parameter τ_2 , multiplied by (-1) , i.e., it is $-\tau_2\tilde{\beta}_2$, where $\tilde{\beta}_2$ is a Gamma distributed random variable with shape parameter α_2 and scale parameter 1.

A Gamma distributed random variable x with shape parameter α and scale parameter 1 has density function

$$h(x; \alpha) = \frac{1}{\Gamma(\alpha)} e^{-x} x^{\alpha-1}$$

and distribution function

$$H(x; \alpha) = \frac{1}{\Gamma(\alpha)} \int_0^x e^{-t} t^{\alpha-1} dt. \quad (3)$$

Functions for computing $\Gamma(\alpha)$ and $H(x; \alpha)$ and the inverse of $H(x; \alpha)$ (for computing quantiles) are widely available. For example, SPSS (SPSS, 1998) can compute $H(x; \alpha)$ and its inverse, Matlab (MathWorks, 1999) can compute $\Gamma(\alpha)$, and further routines can be found in Press, Teukolsky, Vetterling, and Flannery (1992), DiDonato and Morris (1986, 1987), Shea (1988), and Brown, Lovato, and Russell (1994).

The above observations imply that in IV(a) and (b), $r = -(\tau_2/\beta_1)\tilde{r}$, where $\tilde{r} = \tilde{\beta}_2$ is a Gamma distributed random variable with shape parameter α_2 and scale parameter 1. In IV(c), $\beta_1 = -\tau_1\tilde{\beta}_1$ and $\beta_2 = -\tau_2\tilde{\beta}_2$, where $\tilde{\beta}_1$ and $\tilde{\beta}_2$ are independent Gamma distributed random variables with shape parameters α_1 and α_2 , respectively, and scale parameters equal to 1. This implies that $r = (\tau_2/\tau_1)\tilde{r}$, where \tilde{r} has density function

$$g(r; \alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{r^{\alpha_2-1}}{(1+r)^{\alpha_1+\alpha_2}}. \quad (4)$$

Hogg and Klugman (1983) call this the *generalized Pareto* distribution, but other authors (e.g., Davison & Smith, 1990) use this term for another distribution, with density function $\sigma^{-1}(1 - \sigma^{-1}ky)^{1/k-1}$. From (4), it follows that $\tilde{u} = \tilde{r}/(1 + \tilde{r})$ is distributed as a $B(\alpha_2, \alpha_1)$ beta random variable (cf. Johnson & Kotz, 1970, chapter 26). A $B(a, b)$ random variable has distribution function

$$I_x(a, b) = \frac{1}{B(a, b)} \int_0^x t^{a-1} (1-t)^{b-1} dt. \quad (5)$$

Because the transformation from \tilde{r} to \tilde{u} is monotonically increasing, it follows that $G(r) = \Pr(\tilde{r} \leq r) = \Pr(\tilde{u} \leq u) = I_u(\alpha_2, \alpha_1)$, where $u = r/(1 + r)$. Quantiles of \tilde{r} can similarly be obtained from the inverse $K(q; a, b)$, which is defined by the relation

$$I_x(a, b) = q \quad \Longleftrightarrow \quad K(q; a, b) = x. \quad (6)$$

Functions for computing $I_x(a, b)$ and $K(q; a, b)$ are widely available, e.g., in SPSS (SPSS, 1998), Press et al. (1992), DiDonato and Morris (1992), and Brown et al. (1994).

In model IV(d), the dependence between the two Gamma random variables β_1 and β_2 is modeled by using the bivariate Gamma distribution of Moran (1969): Let ζ_1 and ζ_2 be correlated normally distributed random variables with mean zero, variance one, and correlation ρ . Then the two gamma variates $\tilde{\beta}_1$ and $\tilde{\beta}_2$ are assumed to be generated as

$$\tilde{\beta}_1 = H^{-1}[\Phi(\zeta_1); \alpha_1]; \quad (7a)$$

$$\tilde{\beta}_2 = H^{-1}[\Phi(\zeta_2); \alpha_2]; \quad (7b)$$

where $H(\cdot; \cdot)$ is defined in (3) and $\Phi(\cdot)$ denotes the standard normal distribution function. Clearly, $\tilde{\beta}_1$ and $\tilde{\beta}_2$ are dependent and their marginal distributions are Gamma distributions with shape parameters α_1 and α_2 , respectively, and scale parameters equal to 1. The random coefficients β_1 and β_2 are given by $\beta_1 = -\tau_1 \tilde{\beta}_1$ and $\beta_2 = -\tau_2 \tilde{\beta}_2$. Hence, $r = (\tau_2/\tau_1)\tilde{r}$, with $\tilde{r} = \tilde{\beta}_2/\tilde{\beta}_1$. The probability density function of \tilde{r} is

$$g(r) = C(\alpha_1, \alpha_2, \rho) \frac{r^{\alpha_2-1}}{(1+r)^{\alpha_1+\alpha_2}} \int_0^{+\infty} v^{\alpha_1+\alpha_2-1} e^{-v} h(v; r; \alpha_1, \alpha_2, \rho) dv, \quad (8)$$

where

$$C(\alpha_1, \alpha_2, \rho) = \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\sqrt{1-\rho^2}},$$

$$h(v; r; \alpha_1, \alpha_2, \rho) = \exp \left[-\frac{1}{2} \frac{\rho^2 z_2^2 - 2\rho z_1 z_2 + \rho^2 z_1^2}{1-\rho^2} \right],$$

$$z_1 = \Phi^{-1} \left[H \left(\frac{1}{1+r} v; \alpha_1 \right) \right],$$

$$z_2 = \Phi^{-1} \left[H \left(\frac{r}{1+r} v; \alpha_2 \right) \right],$$

$\Phi(\cdot)$ is the standard normal distribution function, and $H(\cdot; \cdot)$ is defined in (3). Consequently, its distribution function is

$$G(r) = \int_0^r g(u) du. \quad (9)$$

The density function (8) is most easily computed using Gauss-Laguerre numerical integration (see, e.g., Press et al., 1992). The distribution function (9) is most easily computed by simulation analogously to the distribution function for models II(c) and II(d): a large number B , say, of drawings of $(\zeta_1, \zeta_2)'$ from the estimated bivariate normal distribution are generated, giving $(\zeta_{1b}, \zeta_{2b})'$, $b = 1, \dots, B$, from which $(\tilde{\beta}_{1b}, \tilde{\beta}_{2b})'$, are obtained by using the formula (7). This gives $\tilde{r}_b = \tilde{\beta}_{2b}/\tilde{\beta}_{1b}$ and $G(r)$.

Analogous to the normal and lognormal models, the model as estimated is a reparameterized version. Instead of estimating the correlation matrix Σ of ζ directly, it is written as $\Sigma = [\text{diag}(LL')]^{-1/2} LL' [\text{diag}(LL')]^{-1/2}$, where L is a lower triangular matrix with unit diagonal elements. Again, this becomes relevant (for variant (d) only) for the computation of the standard errors.

In model V, the distribution of (β_1, β_2) is discrete, with mass points (β_{1j}, β_{2j}) , $j = 1, \dots, J$, and probabilities

$$\pi_j = \Pr[(\beta_1, \beta_2) = (\beta_{1j}, \beta_{2j})].$$

The model as estimated is a reparameterized version with parameters λ_j such that

$$\pi_j = \frac{\exp(\lambda_j)}{\sum_{i=1}^J \exp(\lambda_i)}.$$

As a necessary arbitrary normalization, $\lambda_J = 0$ was taken. In the empirical analysis, $J = 9$.

The density functions of the models II–IV and the probability mass function of model V corresponding to the estimated random coefficient distributions are depicted in the figures in Meijer and Rouwendal (2000, 2004).

3 Characteristics of the VOT distributions

In Meijer and Rouwendal (2000, 2004), we presented four characteristics of the VOT distributions: mean, standard deviation, mode, and median. Here we give analytical and/or computational formulas for these characteristics.

In Table 1, formulas for the means of the VOT distributions are given for the various models. Most of these are trivial, well-known, or given explicitly in Meijer and Rouwendal (2000, 2004) and will therefore not be discussed here. The only exception is the mean for model IV(d). As discussed in section 2, we can write $r = (\tau_2/\tau_1)\tilde{r}$, where the density of \tilde{r} is given in (8). Consequently, the mean of r is $\xi = (\tau_2/\tau_1)\tilde{\xi}$, where $\tilde{\xi}$ is the mean of \tilde{r} ,

$$\tilde{\xi} \equiv \int_0^\infty r g(r) dr. \quad (10)$$

For this model, to prove or disprove existence of the mean is not easy. This issue is discussed at length in section 10. Based on the analysis in that section, we use the rule of thumb that the mean (10) exists whenever $\alpha_1 > 1$. However, if ρ is not close to zero, the threshold value may be different from 1. An approximation of the mean is obtained most easily by generating a large number of drawings from the estimated distribution and computing the sample average of these drawings.

Table 1: Formulas for the means of the VOT.

Model	Mean
I	β_2/β_1
II(a, b)	μ_2/β_1
II(c, d)	does not exist
III(a, b)	$\exp[(\mu_2 - \mu_1) + \frac{1}{2}\sigma_2^2]$
III(c)	$\exp[(\mu_2 - \mu_1) + \frac{1}{2}(\sigma_1^2 + \sigma_2^2)]$
III(d)	$\exp[(\mu_2 - \mu_1) + \frac{1}{2}(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})]$
IV(a, b)	$(-\tau_2/\beta_1)\alpha_2$
IV(c)	$(\tau_2/\tau_1)\alpha_2/(\alpha_1 - 1)$, provided $\alpha_1 > 1$
IV(d)	$(\tau_2/\tau_1)\tilde{\xi}$, with $\tilde{\xi}$ as in (10); provided it exists
V	$\sum_{j=1}^J \pi_j \beta_{2j}/\beta_{1j}$

In Table 2, formulas for the variances of the VOT distributions are given for the various models. Of course, the standard deviations are the (positive) square roots of the variances. Again, most of these formulas are trivial, well-known, or given explicitly in Meijer and Rouwendal (2000, 2004) and will therefore not be discussed here. Here, we

only discuss the variance for model IV(d). The variance of r in model IV(d) is $(\tau_2/\tau_1)^2$ times the variance of \tilde{r} as defined above. The variance of \tilde{r} is

$$\tilde{\omega}^2 \equiv \int_0^\infty r^2 g(r) dr - \tilde{\xi}^2, \quad (11)$$

with $g(r)$ as in (8) and $\tilde{\xi}$ as in (10). The standard deviation $\tilde{\omega}$ of \tilde{r} is the (positive) square root of $\tilde{\omega}^2$. Again, for this model, to prove or disprove existence of the variance (or standard deviation) is not easy. Based on the analysis in section 10, we use the rule of thumb that the variance (and standard deviation) exists whenever $\alpha_1 > 2$. However, if ρ is not close to zero, the threshold value may be different from 2. Computationally, an approximation to the variance (standard deviation) is most easily obtained by using the sample variance (standard deviation) of a large number of simulated drawings from the estimated distribution.

Table 2: Formulas for the variances of the VOT.

Model	Variance
I	0
II(a, b)	σ_2^2/β_1^2
II(c, d)	does not exist
III(a, b)	$\exp[2(\mu_2 - \mu_1) + 2\sigma_2^2] - \exp[2(\mu_2 - \mu_1) + \sigma_2^2]$
III(c)	$\exp[2(\mu_2 - \mu_1) + 2(\sigma_1^2 + \sigma_2^2)] - \exp[2(\mu_2 - \mu_1) + (\sigma_1^2 + \sigma_2^2)]$
III(d)	$\exp[2(\mu_2 - \mu_1) + 2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})]$ $- \exp[2(\mu_2 - \mu_1) + (\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})]$
IV(a, b)	$(-\tau_2/\beta_1)^2 \alpha_2$
IV(c)	$(\tau_2/\tau_1)^2 \frac{\alpha_2(\alpha_1 + \alpha_2 - 1)}{(\alpha_1 - 1)^2(\alpha_1 - 2)}, \quad \text{provided } \alpha_1 > 2$
IV(d)	$(\tau_2/\tau_1)^2 \tilde{\omega}^2, \quad \text{with } \tilde{\omega}^2 \text{ as in (11); provided it exists}$
V	$\sum_{j=1}^J \pi_j (\beta_{2j}/\beta_{1j})^2 - (\text{mean})^2$

In Table 3, formulas for the modes of the VOT distributions are given for the various models. Again, most of these formulas are trivial or can be derived easily from the density formulas. For the models II(c) and II(d), the density function is given in (1), the maximum of which does not have an obvious closed-form solution. Therefore, the mode has to be found by some form of (univariate) numerical maximization, e.g., by a simple Golden Section search (Scales, 1985, chapter 2). It is well-known that the density (1) may be bimodal (Marsaglia, 1965), so in principle, we may have to report

two modes. In fact, the estimated density for model II(d) does have a second mode on the negative part of the real line, but this mode is very small and we report only the global maximum.

For model IV(d), we have no explicit formula for the mode as well, so here we also have to find the maximum numerically. However, as with the mean and variance, for this model, to prove or disprove existence of the mode is not easy. Based on the analysis in section 10, we use the rule of thumb that the mode exists whenever $\alpha_2 > 1$. However, if ρ is not close to zero, the threshold value may be different from 1.

For model V, the mode is simply the estimated VOT of the class with the highest estimated class probability. Strictly speaking, it is possible that two or more classes have the same VOT, but the probability that the estimates of the VOT of two different classes are equal is zero, so we can safely ignore this possibility.

Table 3: Formulas for the modes of the VOT.

Model	Mode
I	β_2/β_1
II(a, b)	μ_2/β_1
II(c, d)	$\arg \max_r g(r)$, with $g(r)$ as in (1)
III(a, b)	$\exp[(\mu_2 - \mu_1) - \sigma_2^2]$
III(c)	$\exp[(\mu_2 - \mu_1) - (\sigma_1^2 + \sigma_2^2)]$
III(d)	$\exp[(\mu_2 - \mu_1) - (\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})]$
IV(a, b)	$(-\tau_2/\beta_1)(\alpha_2 - 1)$, provided $\alpha_2 > 1$
IV(c)	$(\tau_2/\tau_1)(\alpha_2 - 1)/(\alpha_1 + 1)$, provided $\alpha_2 > 1$
IV(d)	$(\tau_2/\tau_1) \arg \max_r g(r)$, with $g(r)$ as in (8); provided it exists
V	(β_{2j}/β_{1j}) , where $j = \arg \max_i \pi_i$

In Table 4, formulas for the medians of the VOT distributions are given for the various models. Again, most of these formulas are trivial or have been discussed in the more general context of quantiles in section 2. For the models II(c), II(d), and IV(d), no closed-form expression for the median exists, nor can it be computed using standard routines for “special functions”. The medians can, however, be easily approximated by generating a large number of drawings from the estimated distributions and computing the sample medians of these drawings.

Table 4: Formulas for the medians of the VOT.

Model	Median
I	β_2/β_1
II(a, b)	μ_2/β_1
II(c, d)	$G^{-1}(1/2)$, with $G(r)$ as in (2)
III(a, b)	$\exp(\mu_2 - \mu_1)$
III(c, d)	$\exp(\mu_2 - \mu_1)$
IV(a, b)	$(-\tau_2/\beta_1)H^{-1}(1/2; \alpha_2)$, with $H(x; \alpha)$ as in (3)
IV(c)	$(\tau_2/\tau_1) \frac{K(1/2; \alpha_2, \alpha_1)}{1 - K(1/2; \alpha_2, \alpha_1)}$, with $K(q; \alpha_2, \alpha_1)$ as in (6)
IV(d)	$(\tau_2/\tau_1)G^{-1}(1/2)$, with $G(r)$ as in (9)
V	(β_{2j}/β_{1j}) , where j is such that $\sum_{i=1}^J \pi_i I(\beta_{2i}/\beta_{1i} < \beta_{2j}/\beta_{1j}) < 1/2$ and $1/2 \leq \sum_{i=1}^J \pi_i I(\beta_{2i}/\beta_{1i} \leq \beta_{2j}/\beta_{1j})$

4 Standard errors

Standard errors of the characteristics that are discussed in section 3 can be obtained by using the delta method. Other ways of obtaining standard errors, most importantly the bootstrap (e.g., Davison & Hinkley, 1997), are hard to implement for our purposes and will therefore not be discussed.

4.1 The delta method

The principle of the delta method is as follows: The estimation process has resulted in an estimated parameter vector $\hat{\theta}$, say, and an estimated covariance matrix $\hat{V} = \hat{A}/N$ of $\hat{\theta}$, where N is the sample size. Under suitable assumptions, $\text{plim } \hat{\theta} = \theta_0$, $\text{plim } \hat{A} = A_0$, and

$$\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, A_0).$$

It follows that standard errors for the elements of $\hat{\theta}$ are given by

$$\text{se}(\hat{\theta}_i) = \sqrt{\hat{A}_{ii}}/\sqrt{N} = \sqrt{\hat{V}_{ii}}.$$

Now, suppose that we are interested in a (possibly) vector-valued function of θ , $\phi = b(\theta)$, say. If $b(\cdot)$ is differentiable in an open neighborhood of θ_0 , the delta method implies that

$$\sqrt{N}(\hat{\phi} - \phi_0) \xrightarrow{d} \mathcal{N}(0, B_0 A_0 B_0'), \quad (12)$$

where $\phi_0 = b(\theta_0)$, $B_0 = B(\theta_0)$, and

$$B(\theta) = \frac{\partial b}{\partial \theta'}(\theta)$$

is the Jacobian matrix with the derivatives of $b(\cdot)$ with respect to its arguments. It follows that standard errors for the elements of ϕ can be computed as

$$\text{se}(\hat{\phi}_j) = \sqrt{(\hat{B} \hat{V} \hat{B}')_{jj}},$$

where $\hat{B} = B(\hat{\theta})$. Note, however, that (12) does not imply that $\lim_{N \rightarrow \infty} E(\hat{\phi}) = \phi_0$ or $\lim_{N \rightarrow \infty} N \text{Var}(\hat{\phi}_j) = (B_0 A_0 B_0')_{jj}$. These expectations may not even exist. It does, however, imply that

$$\lim_{N \rightarrow \infty} \Pr \left[\hat{\phi}_j - z \text{se}(\hat{\phi}_j) \leq \phi_0 \leq \hat{\phi}_j + z \text{se}(\hat{\phi}_j) \right] = 1 - \alpha, \quad (13)$$

where $z = \Phi^{-1}(1 - \frac{1}{2}\alpha)$ and $0 < \alpha \leq \frac{1}{2}$. Therefore, confidence intervals for ϕ_0 can be based on (13). A more extensive discussion of the delta method can be found in Rao (1973, p. 388) or Wansbeek and Meijer (2000, pp. 369–373).

Apparently, all we need to be able to compute standard errors is the Jacobian matrix \hat{B} and, of course, the estimates $\hat{\theta}$ and their estimated covariance matrix \hat{V} , as provided by the initial estimation process. For our empirical problem, the estimation of $\hat{\theta}$ is discussed in Meijer and Rouwendal (2000, 2004), and it has been implemented in GAUSS (Aptech Systems, 1998). The matrices \hat{V} have been computed in GAUSS by the BHHH method (Berndt, Hall, Hall, & Hausman, 1974). In the following, we focus on the computation of \hat{B} . The vector ϕ consists of the mean of the VOT (ξ), its standard deviation (ω), its mode (M), and its median (m), whenever these exist. Hence, we need (computational) formulas for $\partial \xi / \partial \theta$, $\partial \omega / \partial \theta$, $\partial M / \partial \theta$, and $\partial m / \partial \theta$. For the standard deviation, it is sometimes easier to compute the derivatives of the variance, $\partial \omega^2 / \partial \theta$, and use the formula

$$\frac{\partial \omega}{\partial \theta} = \frac{\partial \omega}{\partial \omega^2} \frac{\partial \omega^2}{\partial \theta} = \frac{1}{2\omega} \frac{\partial \omega^2}{\partial \theta}. \quad (14)$$

In most cases, expressions of the derivatives can be obtained straightforwardly from the formulas in the tables given in section 3. In some cases, however, this is not possible and therefore, we will first discuss the general case. After that, we give specific formulas.

4.2 General expressions for the derivatives

Let, as above, the density function of the VOT be denoted by $g(r)$, which is implicitly a function of θ as well. The distribution function is accordingly denoted by $G(r) = \int_{-\infty}^r g(u) du$. The discrete case will be discussed only for our specific situation later.

We start with the mean ξ (assuming it exists),

$$\xi = E(r) = \int_{-\infty}^{+\infty} r g(r) dr.$$

Hence,

$$\begin{aligned} \frac{\partial \xi}{\partial \theta} &= \frac{\partial}{\partial \theta} \int_{-\infty}^{+\infty} r g(r) dr \\ &= \int_{-\infty}^{+\infty} r \frac{\partial g(r)}{\partial \theta} dr \end{aligned} \quad (15a)$$

$$\begin{aligned} &= \int_{-\infty}^{+\infty} r \frac{\partial g(r)}{\partial \theta} \frac{1}{g(r)} g(r) dr \\ &= E \left[r \frac{1}{g(r)} \frac{\partial g(r)}{\partial \theta} \right] \end{aligned} \quad (15b)$$

$$= E \left[r \frac{\partial \log g(r)}{\partial \theta} \right]. \quad (15c)$$

In some cases, (15a) will lead to the most convenient expressions, but in other cases, the derivative may have to be approximated by simulation, in which case the expressions (15b) and (15c) may be convenient.

For the variance ω^2 (if it exists), we have

$$\omega^2 = E(r^2) - \xi^2 = \int_{-\infty}^{+\infty} r^2 g(r) dr - \xi^2,$$

so, analogously to the derivatives of the mean above, we obtain

$$\begin{aligned} \frac{\partial \omega^2}{\partial \theta} &= \int_{-\infty}^{+\infty} r^2 \frac{\partial g(r)}{\partial \theta} dr - 2\xi \frac{\partial \xi}{\partial \theta} \\ &= E \left[r^2 \frac{1}{g(r)} \frac{\partial g(r)}{\partial \theta} \right] - 2\xi \frac{\partial \xi}{\partial \theta} \\ &= E \left[r^2 \frac{\partial \log g(r)}{\partial \theta} \right] - 2\xi \frac{\partial \xi}{\partial \theta}. \end{aligned}$$

The mode M (if it exists) is the value of r for which the density attains a maximum,

$$M = \arg \max_r g(r).$$

Hence, locally, the mode is the value of r that solves the first order equation

$$\frac{\partial g(r)}{\partial r} = 0.$$

By the implicit function theorem (e.g., Wansbeek & Meijer, 2000, p. 371), it follows that

$$\frac{\partial M}{\partial \theta} = - \left(\frac{\partial^2 g}{\partial r^2} \right)^{-1} \left(\frac{\partial^2 g}{\partial r \partial \theta} \right) \Big|_{r=M}. \quad (16)$$

The median m is the value of r for which the distribution function has the value $1/2$, i.e., $m = G^{-1}(1/2)$. Consequently, the median is the value of r that solves the equation

$$G(r) - 1/2 = 0.$$

Thus, applying the implicit function theorem again, it follows that

$$\frac{\partial m}{\partial \theta} = - \left(\frac{\partial G}{\partial r} \right)^{-1} \left(\frac{\partial G}{\partial \theta} \right) \Big|_{r=m} \quad (17a)$$

$$\begin{aligned} &= - \frac{1}{g(m)} \frac{\partial G(m)}{\partial \theta} \\ &= - \frac{1}{g(m)} \frac{\partial}{\partial \theta} \int_{-\infty}^m g(r) \, dr \\ &= - \frac{1}{g(m)} \int_{-\infty}^m \frac{\partial g(r)}{\partial \theta} \, dr \end{aligned} \quad (17b)$$

$$\begin{aligned} &= - \frac{1}{g(m)} \int_{-\infty}^{+\infty} I(r \leq m) \frac{\partial g(r)}{\partial \theta} \, dr \\ &= - \frac{1}{g(m)} \mathbb{E} \left[I(r \leq m) \frac{1}{g(r)} \frac{\partial g(r)}{\partial \theta} \right] \end{aligned} \quad (17c)$$

$$= - \frac{1}{g(m)} \mathbb{E} \left[I(r \leq m) \frac{\partial \log g(r)}{\partial \theta} \right], \quad (17d)$$

where, as usual, $I(\cdot)$ denotes the indicator function. Again, depending on the application, (17b), (17c), or (17d) may be preferred.

Note that in several cases, the characteristic under study has the form

$$\phi_j = (\theta_k / \theta_l) \tilde{\phi}_j,$$

where $\tilde{\phi}_j$ does not depend on θ_k and θ_l . In such cases, the derivatives with respect to θ_k and θ_l are easily obtained as

$$\begin{aligned} \frac{\partial \phi_j}{\partial \theta_k} &= \frac{1}{\theta_k} \phi_j \\ \frac{\partial \phi_j}{\partial \theta_l} &= - \frac{1}{\theta_l} \phi_j. \end{aligned}$$

Thus, once ϕ_j has been obtained, the derivatives with respect to θ_k and θ_l are easily obtained and the derivatives with respect to the remaining parameters are

$$\frac{\partial \phi_j}{\partial \theta_i} = (\theta_k / \theta_l) \frac{\partial \tilde{\phi}_j}{\partial \theta_i}, \quad i \neq k, l.$$

Therefore, only the derivatives of $\tilde{\phi}_j$ with respect to θ_i need to be computed, for $i \neq k, l$.

Finally, in some situations the derivatives are very hard to compute analytically, and it may be preferable to approximate the derivative by a numerical derivative:

$$(\hat{B}_{\text{num}})_{ji} = \frac{b_j(\hat{\theta} + \delta e_i) - b_j(\hat{\theta} - \delta e_i)}{2\delta},$$

where e_i is the i -th unit vector (i.e., the i -th column of the identity matrix), δ is a small number, e.g., $\delta = 10^{-6}$, and $b_j(\cdot)$ is the j -th element of the function $b(\cdot)$. See, e.g., Meijer (1998, pp. 64–65) for a discussion of numerical differentiation.

4.3 Specific expressions for easy cases

In this section, we give formulas for the derivatives of the mean ξ , the standard deviation ω , the mode M , and the median m with respect to the parameters in the original parameter vector θ for the various model specifications considered. These formulas are straightforwardly derived from the expressions of these characteristics in terms of the parameters as given in section 3 and will therefore not be discussed in great detail. Occasionally, we give the formula for the derivatives of the variance ω^2 with respect to θ . The derivatives of the standard deviation with respect to θ then follow immediately from (14). Derivatives that are not straightforward to compute will be discussed in more detail in later sections.

For model I, the standard logit model, the original parameters are β_1 and β_2 , and the relevant derivatives are

$$\frac{\partial(\xi, \omega, M, m)'}{\partial(\beta_1, \beta_2)} = \begin{bmatrix} -\xi/\beta_1 & 1/\beta_1 \\ 0 & 0 \\ -M/\beta_1 & 1/\beta_1 \\ -m/\beta_1 & 1/\beta_1 \end{bmatrix}.$$

For models II(a) and II(b), we have

$$\frac{\partial(\xi, \omega, M, m)'}{\partial(\beta_1, \mu_2, \sigma_2^2)} = \begin{bmatrix} -\xi/\beta_1 & 1/\beta_1 & 0 \\ -\omega/\beta_1 & 0 & \omega/(2\sigma_2^2) \\ -M/\beta_1 & 1/\beta_1 & 0 \\ -m/\beta_1 & 1/\beta_1 & 0 \end{bmatrix}.$$

To obtain the derivatives with respect to the estimated parameters, we need to postmultiply this matrix by the Jacobian matrix $\partial(\beta_1, \mu_2, \sigma_2^2)' / \partial\theta'$. Apart from trivial zeros and ones, this contains the derivatives of σ_2^2 with respect to the elements of the Cholesky root L . These have the same formulas as for model III, which is discussed below because it contains more general formulas.

For models II(c) and II(d), i.e., normally distributed coefficients with a random cost coefficient, the mean, variance, and standard deviation do not exist, and thus their derivatives (and standard errors) do not exist as well. The derivatives of the mode and median are complicated and will be discussed in section 5.

Lognormal (III)

For the models III(a)–III(d), the relevant derivatives are

$$\frac{\partial(\xi, \omega^2, M, m)'}{\partial(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_{12})} = \begin{bmatrix} -\xi & \xi & \xi/2 & \xi/2 & -\xi \\ -2\omega^2 & 2\omega^2 & E_1 & E_1 & -2E_1 \\ -M & M & -M & -M & 2M \\ -m & m & 0 & 0 & 0 \end{bmatrix},$$

where $E_1 = 2E_2 - E_3$, $E_2 = \exp[2(\mu_2 - \mu_1) + 2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})]$, and $E_3 = \exp[2(\mu_2 - \mu_1) + (\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})]$, and the relevant columns are removed in the variants for which the corresponding parameter is fixed to zero.

As already indicated in the discussion of model II above, this matrix must be postmultiplied by the matrix of derivatives of these parameters with respect to the estimated parameters, which are the μ 's of all coefficients and the (free) elements L_{kl} , $k \geq l$, of the Cholesky root of the covariance matrix Σ of all coefficients. The nontrivial part of this consists of the derivatives $\partial\Sigma_{ij} / \partial L_{kl}$. What we have called σ_1^2 up till now is Σ_{ii} for some i , and similarly $\sigma_2^2 = \Sigma_{jj}$ and $\sigma_{12} = \Sigma_{ij}$. Then, from $\Sigma = LL'$, it follows that

$$\sigma_1^2 = \Sigma_{ii} = \sum_{k=1}^i L_{ik}^2,$$

from which we obtain

$$\frac{\partial\sigma_1^2}{\partial L_{ik}} = 2L_{ik}, \quad k = 1, \dots, i. \quad (18a)$$

Analogously,

$$\frac{\partial \sigma_2^2}{\partial L_{jk}} = 2L_{jk}, \quad k = 1, \dots, j; \quad (18b)$$

$$\frac{\partial \sigma_{12}}{\partial L_{ik}} = L_{jk}, \quad k = 1, \dots, \min(i, j); \quad (18c)$$

$$\frac{\partial \sigma_{12}}{\partial L_{jk}} = L_{ik}, \quad k = 1, \dots, \min(i, j). \quad (18d)$$

Finally, for the model II(d) later on, we will need the derivative of the correlation coefficient with respect to the parameters. By the chain rule, we need te above, plus the derivatives of ρ with respect to σ_1^2 , σ_2^2 , and σ_{12} . From $\rho = \sigma_{12}(\sigma_1^2)^{-1/2}(\sigma_2^2)^{-1/2}$, we obtain

$$\frac{\partial \rho}{\partial \sigma_{12}} = (\sigma_1^2)^{-1/2}(\sigma_2^2)^{-1/2}; \quad (19a)$$

$$\frac{\partial \rho}{\partial \sigma_1^2} = -\rho/(2\sigma_1^2); \quad (19b)$$

$$\frac{\partial \rho}{\partial \sigma_2^2} = -\rho/(2\sigma_2^2). \quad (19c)$$

Obvuiosly, formulas (18) and (19) are only relevant for the models in which the corresponding parameters are well-defined and not restricted to zero.

Gamma (IV)

For the Gamma models IV(a)–IV(d), note that, instead of the scale parameter τ , its negative ($-\tau$) has been estimated whenever applicable. Hence, for the models IV(a) and IV(b) with a nonrandom cost coefficient, the original parameters are $(\beta_1, -\tau_2, \alpha_2)$, and the relevant derivatives are

$$\frac{\partial(\xi, \omega, M, m)'}{\partial(\beta_1, -\tau_2, \alpha_2)} = \begin{bmatrix} -\xi/\beta_1 & \xi/(-\tau_2) & \xi/\alpha_2 \\ -\omega/\beta_1 & \omega/(-\tau_2) & \omega/(2\alpha_2) \\ -M/\beta_1 & M/(-\tau_2) & M/(\alpha_2 - 1) \\ -m/\beta_1 & m/(-\tau_2) & * \end{bmatrix},$$

provided the relevant characteristic exists. In this case, this may only be a problem for the mode, which exists only if $\alpha_2 > 1$. The asterisk in the lower right element means a more difficult expression. We will use this notation throughout this section. In this

case, the expression is

$$\frac{\partial m}{\partial \alpha_2} = - \frac{(-\tau_2)/\beta_1}{\frac{1}{\Gamma(\alpha_2)} e^{-\tilde{m}} \tilde{m}^{\alpha_2-1}} \frac{\partial H(\tilde{m}; \alpha_2)}{\partial \alpha_2} = -m \Gamma(\alpha_2) e^{\tilde{m}} \tilde{m}^{-\alpha_2} \frac{\partial H(\tilde{m}; \alpha_2)}{\partial \alpha_2},$$

where \tilde{m} is implicitly defined by $m = (-\tau_2/\beta_1)\tilde{m}$, and with $H(\cdot; \cdot)$ defined in (3). A function for computing the derivative of $H(x; \alpha)$ with respect to α is given by Moore (1982).

For the model IV(c), with independent Gamma distributed coefficients and a random cost coefficient, the original parameters are $(-\tau_1, -\tau_2, \alpha_1, \alpha_2)$, and the relevant derivatives are

$$\begin{aligned} \frac{\partial(\xi, \omega, M, m)'}{\partial(-\tau_1, -\tau_2)} &= \begin{bmatrix} -\xi/(-\tau_1) & \xi/(-\tau_2) \\ -\omega/(-\tau_1) & \omega/(-\tau_2) \\ -M/(-\tau_1) & M/(-\tau_2) \\ -m/(-\tau_1) & m/(-\tau_2) \end{bmatrix} \\ \frac{\partial(\xi, \omega, M, m)'}{\partial(\alpha_1, \alpha_2)} &= \begin{bmatrix} -\xi/(\alpha_1 - 1) & \xi/\alpha_2 \\ * & * \\ -M/(\alpha_1 + 1) & M/(\alpha_2 - 1) \\ * & * \end{bmatrix}, \end{aligned}$$

provided the relevant characteristic exists. In this case, the mean exists if $\alpha_1 > 1$, the standard deviation exists if $\alpha_1 > 2$, and the mode exists if $\alpha_2 > 1$. The derivatives of the variance with respect to the shape parameters are

$$\begin{aligned} \frac{\partial \omega^2}{\partial \alpha_1} &= \frac{(\tau_2/\tau_1)^2 \alpha_2}{(\alpha_1 - 1)^2 (\alpha_1 - 2)} \left[1 - \frac{(\alpha_1 + \alpha_2 - 1)(3\alpha_1 - 5)}{(\alpha_1 - 1)(\alpha_1 - 2)} \right] \\ \frac{\partial \omega^2}{\partial \alpha_2} &= \frac{(\tau_2/\tau_1)^2 (\alpha_1 + 2\alpha_2 - 1)}{(\alpha_1 - 1)^2 (\alpha_1 - 2)} \end{aligned}$$

The derivatives of the median with respect to the shape parameters are

$$\frac{\partial m}{\partial(\alpha_1, \alpha_2)} = - \frac{(\tau_2/\tau_1) \Gamma(\alpha_1) \Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2) x^{\alpha_2-1} (1-x)^{\alpha_1+1}} \frac{\partial I_x(\alpha_2, \alpha_1)}{\partial(\alpha_1, \alpha_2)},$$

where $x = K(1/2; \alpha_2, \alpha_1)$, $K(\cdot; \cdot, \cdot)$ is defined in (6), and $I_x(\alpha_2, \alpha_1)$ denotes the incomplete Beta function ratio (5). The derivatives of this function with respect to its parameters α_1 and α_2 will be discussed in section 6.

For model IV(d), with dependent Gamma distributed coefficients, the derivatives with respect to $(-\tau_1)$ and $(-\tau_2)$ are straightforward, but the derivatives with respect to

$(\alpha_1, \alpha_2, \rho)$ are complicated, so that we have

$$\frac{\partial(\xi, \omega, M, m)'}{\partial(-\tau_1, -\tau_2, \alpha_1, \alpha_2, \rho)} = \begin{bmatrix} -\xi/(-\tau_1) & \xi/(-\tau_2) & * & * & * \\ -\omega/(-\tau_1) & \omega/(-\tau_2) & * & * & * \\ -M/(-\tau_1) & M/(-\tau_2) & * & * & * \\ -m/(-\tau_1) & m/(-\tau_2) & * & * & * \end{bmatrix}.$$

The derivatives with respect to α_1, α_2 , and ρ are discussed in section 7.

Latent class (V)

For the latent class (discrete) model V, the original parameters are $(\beta_{1j}, \beta_{2j}, \lambda_j)$, $j = 1, \dots, J$, except for the last class (J), for which $\lambda_J = 0$ is a fixed constant. Note that $\partial\pi_j/\partial\lambda_j = \pi_j(1 - \pi_j)$. Hence, the derivatives of the mean are

$$\frac{\partial\xi}{\partial(\beta_{1j}, \beta_{2j}, \lambda_j)} = (-\pi_j\beta_{2j}/\beta_{1j}^2, \pi_j/\beta_{1j}, \pi_j(1 - \pi_j)\beta_{2j}/\beta_{1j}).$$

Of course, for $j = J$, the last element does not exist because $\lambda_J = 0$ by definition. Analogously, the derivatives of the variance are

$$\begin{aligned} \frac{\partial\omega^2}{\partial\beta_{1j}} &= -2\pi_j\frac{\beta_{2j}}{\beta_{1j}^2} \left[\frac{\beta_{2j}}{\beta_{1j}} - \xi \right] \\ \frac{\partial\omega^2}{\partial\beta_{2j}} &= 2\pi_j\frac{1}{\beta_{1j}} \left[\frac{\beta_{2j}}{\beta_{1j}} - \xi \right] \\ \frac{\partial\omega^2}{\partial\lambda_j} &= \pi_j(1 - \pi_j)\frac{\beta_{2j}}{\beta_{1j}} \left[\frac{\beta_{2j}}{\beta_{1j}} - 2\xi \right]. \end{aligned}$$

Again, for $j = J$, the last element does not exist. The derivatives of the mode are

$$\begin{aligned} \frac{\partial M}{\partial\beta_{1j}} &= -M/\beta_{1j} \\ \frac{\partial M}{\partial\beta_{2j}} &= 1/\beta_{1j}, \end{aligned}$$

where $j = \arg \max_i (\beta_{2j}/\beta_{1j})$, i.e., the class number of the mode. All other derivatives are zero, including the derivative with respect to λ_j . Similarly, the derivatives of the median are

$$\begin{aligned} \frac{\partial m}{\partial\beta_{1j}} &= -m/\beta_{1j} \\ \frac{\partial m}{\partial\beta_{2j}} &= 1/\beta_{1j}, \end{aligned}$$

where j is such that

$$\sum_{i=1}^J \pi_i I(\beta_{2i}/\beta_{1i} < \beta_{2j}/\beta_{1j}) < 1/2 \leq \sum_{i=1}^J \pi_i I(\beta_{2i}/\beta_{1i} \leq \beta_{2j}/\beta_{1j}),$$

i.e., the class number of the median. All other derivatives are zero, including the derivative with respect to λ_j .

5 Derivatives for models II(c) and II(d)

For models II(c) and II(d), i.e., with normally distributed coefficients and a random cost coefficient, the mean and standard deviation of the VOT do not exist. Thus, we only have to consider the derivatives of the mode and median with respect to the original parameters. We will derive expressions for the most general model II(d). The corresponding expressions for model II(c) are then obtained by putting $\rho = 0$ and deleting the derivatives with respect to ρ . Note that the derivatives obtained here must be postmultiplied by the expressions given in (18) and (19) above to obtain the derivatives with respect to the estimated parameters.

For the mode, we can use expression (16). Because we have a closed form expression of the density $g(r)$, given in (1), this is straightforward in principle. The derivations are, however, long and tedious. We will give the expressions by successively breaking down the expressions in terms of sub-expressions. The density function can be written as

$$g(r) = E_1 E_2 + E_3 P_2, \quad (20)$$

where

$$\begin{aligned} E_1 &= \left(\frac{1}{\pi}\right) \left(\frac{\sqrt{1-\rho^2}}{\sigma_1 \sigma_2}\right) \left(\frac{1}{a^2(r)}\right) = K_1 C_1 A_0 \\ E_2 &= \exp\left(-\frac{c}{2(1-\rho^2)}\right) \\ E_3 &= \left(\frac{1}{\sqrt{2\pi}}\right) b(r) d(r) \left(\frac{1}{\sigma_1 \sigma_2 a^3(r)}\right) = K_2 b(r) d(r) F_1 \\ P_2 &= \Phi\left(\frac{b(r)}{a(r)\sqrt{1-\rho^2}}\right) - \Phi\left(-\frac{b(r)}{a(r)\sqrt{1-\rho^2}}\right) \\ &= \Phi(Q_1) - \Phi(-Q_1) \\ &= 2\Phi(Q_1) - 1 \\ &= 2P_1 - 1, \end{aligned}$$

with $K_1, K_2, C_1, A_0, F_1, P_1$, and Q_1 implicitly defined. For future use, note that

$$d(r) = \exp(\frac{1}{2}Q_1^2)E_2.$$

Finally, we use the basic expressions given earlier,

$$\begin{aligned} a^2(r) &= \frac{r^2}{\sigma_2^2} - \frac{2\rho r}{\sigma_1\sigma_2} + \frac{1}{\sigma_1^2} \\ b(r) &= \frac{\mu_2 r}{\sigma_2^2} - \frac{\rho(\mu_2 + \mu_1 r)}{\sigma_1\sigma_2} + \frac{\mu_1}{\sigma_1^2} \\ c &= \frac{\mu_2^2}{\sigma_2^2} - \frac{2\rho\mu_1\mu_2}{\sigma_1\sigma_2} + \frac{\mu_1^2}{\sigma_1^2}. \end{aligned}$$

These expressions are in terms of r and the parameters $\tilde{\theta} = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)'$. However, as discussed above, instead of the variances and correlations, the Cholesky root L of the covariance matrix has been estimated. Therefore, the derivatives with respect to $\tilde{\theta}$ must be postmultiplied by the expressions given in (18) and (19) above to obtain the derivatives with respect to the estimated parameters θ .

The first derivative w.r.t. r

The first derivative of the density $g(r)$ with respect to r is

$$\begin{aligned} \frac{\partial g(r)}{\partial r} &= \frac{\partial E_1}{\partial r} E_2 + E_1 \frac{\partial E_2}{\partial r} + \frac{\partial E_3}{\partial r} P_2 + E_3 \frac{\partial P_2}{\partial r} \\ &= \frac{\partial E_1}{\partial r} E_2 + F_3 P_2 + 2E_3 \frac{\partial P_1}{\partial r}, \end{aligned} \tag{21}$$

where $\partial E_2/\partial r = 0$, because E_2 does not depend on r , has been used, and $F_3 \equiv \partial E_3/\partial r$. The derivatives in (21) are

$$\begin{aligned} \frac{\partial E_1}{\partial r} &= K_1 C_1 \frac{\partial A_0}{\partial r} \\ &= K_1 C_1 \left(-\frac{1}{a^4(r)} \frac{\partial a^2(r)}{\partial r} \right) \\ &= K_1 C_1 \frac{1}{a^2(r)} \left(-\frac{1}{a^2(r)} \frac{\partial a^2(r)}{\partial r} \right) \\ &= K_1 C_1 A_0 A_1 \\ &= E_1 A_1 \end{aligned}$$

$$F_3 \equiv \frac{\partial E_3}{\partial r} = K_2 \frac{\partial b(r)}{\partial r} d(r) F_1 + K_2 b(r) \frac{\partial d(r)}{\partial r} F_1 + K_2 b(r) d(r) \frac{\partial F_1}{\partial r}$$

$$\frac{\partial P_1}{\partial r} = \phi(Q_1) \frac{\partial Q_1}{\partial r} = K_2 \exp(-\frac{1}{2} Q_1^2) Q_2,$$

with A_1 and Q_2 implicitly defined. The first derivatives of $a^2(r)$, $b(r)$, and $d(r)$ with respect to r are

$$\frac{\partial a^2(r)}{\partial r} = \frac{2r}{\sigma_2^2} - \frac{2\rho}{\sigma_1 \sigma_2}$$

$$\frac{\partial b(r)}{\partial r} = \frac{\mu_2}{\sigma_2^2} - \frac{\rho \mu_1}{\sigma_1 \sigma_2}$$

$$\frac{\partial d(r)}{\partial r} = Q_1 \exp(\frac{1}{2} Q_1^2) \frac{\partial Q_1}{\partial r} E_2 = d(r) Q_1 Q_2.$$

Consequently,

$$A_1 \equiv -\frac{1}{a^2(r)} \frac{\partial a^2(r)}{\partial r} = -A_0 \left(\frac{2r}{\sigma_2^2} - \frac{2\rho}{\sigma_1 \sigma_2} \right)$$

$$\frac{\partial F_1}{\partial r} = \frac{1}{\sigma_1 \sigma_2} \left(-\frac{3}{2} \frac{1}{a^5(r)} \frac{\partial a^2(r)}{\partial r} \right) = \frac{3}{2} F_1 A_1$$

$$Q_2 \equiv \frac{\partial Q_1}{\partial r} = \frac{1}{a(r) \sqrt{1-\rho^2}} \frac{\partial b(r)}{\partial r} - \frac{1}{2} \frac{b(r)}{a^3(r) \sqrt{1-\rho^2}} \frac{\partial a^2(r)}{\partial r}$$

$$= \frac{1}{a^3(r) \sqrt{1-\rho^2}} \left(a^2(r) \frac{\partial b(r)}{\partial r} - \frac{1}{2} b(r) \frac{\partial a^2(r)}{\partial r} \right)$$

$$= \frac{1}{a^3(r) \sqrt{1-\rho^2}} \left(\frac{1-\rho^2}{\sigma_1^2 \sigma_2^2} (\mu_2 - \mu_1 r) \right)$$

$$= C_1 F_1 (\mu_2 - \mu_1 r).$$

Hence, the expression for F_3 becomes

$$F_3 = K_2 d(r) F_1 \frac{\partial b(r)}{\partial r} + K_2 b(r) d(r) Q_1 Q_2 F_1 + K_2 b(r) d(r) \frac{3}{2} F_1 A_1$$

$$= K_2 d(r) B_1 + E_3 (Q_1 Q_2 + \frac{3}{2} A_1),$$

where $B_1 \equiv F_1 \partial b(r) / \partial r$. Combining terms, we have

$$\frac{\partial g(r)}{\partial r} = A_1 E_1 E_2 + F_3 P_2 + 2E_3 K_2 \exp(-\frac{1}{2} Q_1^2) Q_2$$

$$= A_1 E_1 E_2 + F_3 P_2 + 2K_2 b(r) d(r) F_1 K_2 \exp(-\frac{1}{2} Q_1^2) Q_2$$

$$= A_1 E_1 E_2 + F_3 P_2 + K_1 b(r) E_2 F_1 Q_2.$$

The second derivative w.r.t. r

Continuing in the same way, we find that the second derivative of $g(r)$ with respect to r is

$$\begin{aligned}\frac{\partial^2 g(r)}{\partial r^2} &= \frac{\partial A_1}{\partial r} E_1 E_2 + A_1 \frac{\partial E_1}{\partial r} E_2 + A_1 E_1 \frac{\partial E_2}{\partial r} + \frac{\partial F_3}{\partial r} P_2 + F_3 \frac{\partial P_2}{\partial r} + K_1 \frac{\partial b(r)}{\partial r} E_2 F_1 Q_2 \\ &\quad + K_1 b(r) \frac{\partial E_2}{\partial r} F_1 Q_2 + K_1 b(r) E_2 \frac{\partial F_1}{\partial r} Q_2 + K_1 b(r) E_2 F_1 \frac{\partial Q_2}{\partial r} \\ &= \frac{\partial A_1}{\partial r} E_1 E_2 + A_1^2 E_1 E_2 + \frac{\partial F_3}{\partial r} P_2 + 2K_2 \exp(-\frac{1}{2}Q_1^2) F_3 Q_2 + K_1 B_1 E_2 Q_2 \\ &\quad + \frac{3}{2} K_1 b(r) E_2 F_1 A_1 Q_2 + K_1 b(r) E_2 F_1 \frac{\partial Q_2}{\partial r},\end{aligned}$$

where we have used some results derived previously. Unknown elements of this expression are

$$\begin{aligned}\frac{\partial A_1}{\partial r} &= -\frac{\partial A_0}{\partial r} \left(\frac{2r}{\sigma_2^2} - \frac{2\rho}{\sigma_1 \sigma_2} \right) - \frac{2A_0}{\sigma_2^2} \\ &= -A_0 A_1 \left(\frac{2r}{\sigma_2^2} - \frac{2\rho}{\sigma_1 \sigma_2} \right) - \frac{2A_0}{\sigma_2^2} \\ &= A_1^2 - 2A_0/\sigma_2^2 \\ F_4 &\equiv \frac{\partial F_3}{\partial r} \\ &= K_2 \frac{\partial d(r)}{\partial r} B_1 + K_2 d(r) \frac{\partial B_1}{\partial r} + \frac{\partial E_3}{\partial r} (Q_1 Q_2 + \frac{3}{2} A_1) \\ &\quad + E_3 \left(\frac{\partial Q_1}{\partial r} Q_2 + Q_1 \frac{\partial Q_2}{\partial r} + \frac{3}{2} \frac{\partial A_1}{\partial r} \right) \\ &= K_2 d(r) B_1 Q_1 Q_2 + K_2 d(r) \frac{\partial B_1}{\partial r} + F_3 (Q_1 Q_2 + \frac{3}{2} A_1) \\ &\quad + E_3 \left(Q_2^2 + Q_1 \frac{\partial Q_2}{\partial r} + \frac{3}{2} A_1^2 - 3A_0/\sigma_2^2 \right) \\ \frac{\partial Q_2}{\partial r} &= \frac{\partial C_1}{\partial r} F_1 (\mu_2 - \mu_1 r) + C_1 \frac{\partial F_1}{\partial r} (\mu_2 - \mu_1 r) - \mu_1 C_1 F_1 \\ &= \frac{3}{2} A_1 C_1 F_1 (\mu_2 - \mu_1 r) - \mu_1 C_1 F_1 \\ &= \frac{3}{2} A_1 Q_2 - C_1 F_1 \mu_1,\end{aligned}$$

and

$$\frac{\partial B_1}{\partial r} = \frac{\partial F_1}{\partial r} \frac{\partial b(r)}{\partial r} + F_1 \frac{\partial^2 b(r)}{\partial r^2} = \frac{3}{2} A_1 F_1 \frac{\partial b(r)}{\partial r} = \frac{3}{2} A_1 B_1,$$

so that

$$\begin{aligned}
F_4 &= K_2 d(r) B_1 Q_1 Q_2 + \frac{3}{2} K_2 d(r) A_1 B_1 + F_3 (Q_1 Q_2 + \frac{3}{2} A_1) \\
&\quad + E_3 (Q_2^2 + \frac{3}{2} A_1 Q_1 Q_2 - C_1 F_1 Q_1 \mu_1 + \frac{3}{2} A_1^2 - 3 A_0 / \sigma_2^2) \\
&= (K_2 d(r) B_1 + F_3) (Q_1 Q_2 + \frac{3}{2} A_1) \\
&\quad + E_3 (\frac{3}{2} A_1 (Q_1 Q_2 + A_1) + Q_2^2 - C_1 F_1 Q_1 \mu_1 - 3 A_0 / \sigma_2^2).
\end{aligned}$$

Consequently,

$$\begin{aligned}
\frac{\partial^2 g(r)}{\partial r^2} &= (A_1^2 - 2 A_0 / \sigma_2^2) E_1 E_2 + A_1^2 E_1 E_2 + F_4 P_2 + 2 K_2 \exp(-\frac{1}{2} Q_1^2) F_3 Q_2 \\
&\quad + K_1 B_1 E_2 Q_2 + \frac{3}{2} K_1 b(r) E_2 F_1 A_1 Q_2 + K_1 b(r) E_2 F_1 (\frac{3}{2} A_1 Q_2 - C_1 F_1 \mu_1) \\
&= 2(A_1^2 - A_0 / \sigma_2^2) E_1 E_2 + F_4 P_2 + 2 K_2 \exp(-\frac{1}{2} Q_1^2) F_3 Q_2 \\
&\quad + K_1 E_2 (B_1 Q_2 + b(r) F_1 (3 A_1 Q_2 - C_1 F_1 \mu_1)).
\end{aligned} \tag{22}$$

The first derivatives w.r.t. $\tilde{\theta}$

To compute the derivatives of the mode with respect to the original parameters, we need not only the second derivatives with respect to r , but also the second cross derivatives with respect to r and θ . In order to do so, we need some first order partial derivatives with respect too $\tilde{\theta}$ first. It will turn out that these will also be needed for the derivatives of the median.

From (20), it follows that the derivatives of the density $g(r)$ with respect to the parameters $\tilde{\theta}$ are

$$\frac{\partial g(r)}{\partial \tilde{\theta}'} = E_2 \frac{\partial E_1}{\partial \tilde{\theta}'} + E_1 \frac{\partial E_2}{\partial \tilde{\theta}'} + P_2 \frac{\partial E_3}{\partial \tilde{\theta}'} + 2 E_3 \frac{\partial P_1}{\partial \tilde{\theta}'}. \tag{23}$$

Expressions for the derivatives in this formula are

$$\begin{aligned}
\frac{\partial E_1}{\partial \tilde{\theta}'} &= K_1 A_0 \frac{\partial C_1}{\partial \tilde{\theta}'} + K_1 C_1 \frac{\partial A_0}{\partial \tilde{\theta}'} \\
\frac{\partial E_2}{\partial \tilde{\theta}'} &= -E_2 \left[\frac{1}{2(1-\rho^2)} \frac{\partial c}{\partial \tilde{\theta}'} + \frac{\rho c}{(1-\rho^2)^2} e'_5 \right] = -E_2 D_2 \\
\frac{\partial E_3}{\partial \tilde{\theta}'} &= K_2 \frac{\partial b(r)}{\partial \tilde{\theta}'} d(r) F_1 + K_2 b(r) \frac{\partial d(r)}{\partial \tilde{\theta}'} F_1 + K_2 b(r) d(r) \frac{\partial F_1}{\partial \tilde{\theta}'} \\
\frac{\partial P_1}{\partial \tilde{\theta}'} &= K_2 \exp(-\frac{1}{2} Q_1^2) \frac{\partial Q_1}{\partial \tilde{\theta}'},
\end{aligned}$$

where $e'_5 \equiv (0, 0, 0, 0, 1)$ and D_2 is implicitly defined. Furthermore,

$$\begin{aligned}
\frac{\partial C_1}{\partial \tilde{\theta}'} &= \left(0, 0, -\frac{1}{2}C_1/\sigma_1^2, -\frac{1}{2}C_1/\sigma_2^2, -\frac{\rho}{\sigma_1\sigma_2\sqrt{1-\rho^2}} \right) \\
&= -C_1 \left(0, 0, \frac{1}{2\sigma_1^2}, \frac{1}{2\sigma_2^2}, \frac{\rho}{1-\rho^2} \right) \\
&= -C_1 D_3 \\
\frac{\partial A_0}{\partial \tilde{\theta}'} &= -A_0^2 \frac{\partial a^2(r)}{\partial \tilde{\theta}'} \\
\frac{\partial F_1}{\partial \tilde{\theta}'} &= \left(0, 0, -\frac{1}{2}F_1/\sigma_1^2, -\frac{1}{2}F_1/\sigma_2^2, 0 \right) + \frac{3}{2} \frac{A_0^{1/2}}{\sigma_1\sigma_2} \frac{\partial A_0}{\partial \tilde{\theta}'} \\
&= -\frac{1}{2}F_1 \left[3A_0 \frac{\partial a^2(r)}{\partial \tilde{\theta}'} + \left(0, 0, 1/\sigma_1^2, 1/\sigma_2^2, 0 \right) \right] \\
&= -\frac{1}{2}F_1 D_4 \\
\frac{\partial Q_1}{\partial \tilde{\theta}'} &= \frac{1}{a(r)\sqrt{1-\rho^2}} \frac{\partial b(r)}{\partial \tilde{\theta}'} - \frac{1}{2} \frac{b(r)}{a^3(r)\sqrt{1-\rho^2}} \frac{\partial a^2(r)}{\partial \tilde{\theta}'} + \frac{\rho b(r)}{a(r)(1-\rho^2)^{3/2}} e'_5 \\
&= \frac{1}{a(r)\sqrt{1-\rho^2}} \frac{\partial b(r)}{\partial \tilde{\theta}'} - \frac{1}{2} Q_1 A_0 \frac{\partial a^2(r)}{\partial \tilde{\theta}'} + Q_1 \frac{\rho}{1-\rho^2} e'_5 \\
\frac{\partial a^2(r)}{\partial \tilde{\theta}'} &= \left[0, 0, \frac{1}{\sigma_1^2} \left(\frac{\rho r}{\sigma_1\sigma_2} - \frac{1}{\sigma_1^2} \right), \frac{1}{\sigma_2^2} \left(\frac{\rho r}{\sigma_1\sigma_2} - \frac{r^2}{\sigma_2^2} \right), -\frac{2r}{\sigma_1\sigma_2} \right] \\
\frac{\partial b(r)}{\partial \tilde{\theta}'} &= \left[\frac{1}{\sigma_1^2} - \frac{\rho r}{\sigma_1\sigma_2}, \frac{r}{\sigma_2^2} - \frac{\rho}{\sigma_1\sigma_2}, \frac{1}{\sigma_1^2} \left(\frac{\rho(\mu_2 + \mu_1 r)}{2\sigma_1\sigma_2} - \frac{\mu_1}{\sigma_1^2} \right), \right. \\
&\quad \left. \frac{1}{\sigma_2^2} \left(\frac{\rho(\mu_2 + \mu_1 r)}{2\sigma_1\sigma_2} - \frac{\mu_2 r}{\sigma_2^2} \right), -\frac{\mu_2 + \mu_1 r}{\sigma_1\sigma_2} \right] \\
\frac{\partial c}{\partial \tilde{\theta}'} &= \left[\frac{2\mu_1}{\sigma_1^2} - \frac{2\rho\mu_2}{\sigma_1\sigma_2}, \frac{2\mu_2}{\sigma_2^2} - \frac{2\rho\mu_1}{\sigma_1\sigma_2}, \frac{1}{\sigma_1^2} \left(\frac{\rho\mu_1\mu_2}{\sigma_1\sigma_2} - \frac{\mu_1^2}{\sigma_1^2} \right), \right. \\
&\quad \left. \frac{1}{\sigma_2^2} \left(\frac{\rho\mu_1\mu_2}{\sigma_1\sigma_2} - \frac{\mu_2^2}{\sigma_2^2} \right), -\frac{2\mu_1\mu_2}{\sigma_1\sigma_2} \right] \\
\frac{\partial d(r)}{\partial \tilde{\theta}'} &= d(r) Q_1 \frac{\partial Q_1}{\partial \tilde{\theta}'} + \exp(\frac{1}{2}Q_1^2) \frac{\partial E_2}{\partial \tilde{\theta}'} \\
&= d(r) \left(Q_1 \frac{\partial Q_1}{\partial \tilde{\theta}'} - D_2 \right),
\end{aligned}$$

in which D_3 and D_4 are implicitly defined. Using these results, more concise

expressions for the derivatives of E_1 and E_3 can be obtained:

$$\begin{aligned}
\frac{\partial E_1}{\partial \tilde{\theta}'} &= -K_1 C_1 D_3 A_0 - K_1 C_1 A_0^2 \frac{\partial a^2(r)}{\partial \tilde{\theta}'} \\
&= -E_1 \left(D_3 + A_0 \frac{\partial a^2(r)}{\partial \tilde{\theta}'} \right) \\
\frac{\partial E_3}{\partial \tilde{\theta}'} &= K_2 d(r) F_1 \frac{\partial b(r)}{\partial \tilde{\theta}'} + K_2 b(r) d(r) F_1 \left(Q_1 \frac{\partial Q_1}{\partial \tilde{\theta}'} - D_2 \right) + K_2 b(r) d(r) \left(-\frac{1}{2} F_1 D_4 \right) \\
&= K_2 d(r) F_1 \frac{\partial b(r)}{\partial \tilde{\theta}'} + E_3 \left(Q_1 \frac{\partial Q_1}{\partial \tilde{\theta}'} - D_2 - \frac{1}{2} D_4 \right),
\end{aligned}$$

which can be used in (23) to evaluate the first differences of the density with respect to $\tilde{\theta}'$.

The second cross derivatives

Starting from (23), we find that the second cross partial derivatives are

$$\begin{aligned}
\frac{\partial^2 g(r)}{\partial r \partial \tilde{\theta}'} &= \frac{\partial E_2}{\partial r} \frac{\partial E_1}{\partial \tilde{\theta}'} + E_2 \frac{\partial^2 E_1}{\partial r \partial \tilde{\theta}'} + \frac{\partial E_1}{\partial r} \frac{\partial E_2}{\partial \tilde{\theta}'} + E_1 \frac{\partial^2 E_2}{\partial r \partial \tilde{\theta}'} + \frac{\partial P_2}{\partial r} \frac{\partial E_3}{\partial \tilde{\theta}'} + P_2 \frac{\partial^2 E_3}{\partial r \partial \tilde{\theta}'} \\
&\quad + 2 \frac{\partial E_3}{\partial r} \frac{\partial P_1}{\partial \tilde{\theta}'} + 2 E_3 \frac{\partial^2 P_1}{\partial r \partial \tilde{\theta}'} \\
&= E_2 \frac{\partial^2 E_1}{\partial r \partial \tilde{\theta}'} - A_1 E_1 E_2 D_2 + 2 K_2 \exp(-\frac{1}{2} Q_1^2) Q_2 \frac{\partial E_3}{\partial \tilde{\theta}'} + P_2 \frac{\partial^2 E_3}{\partial r \partial \tilde{\theta}'} \\
&\quad + 2 F_3 K_2 \exp(-\frac{1}{2} Q_1^2) \frac{\partial Q_1}{\partial \tilde{\theta}'} + 2 E_3 \frac{\partial^2 P_1}{\partial r \partial \tilde{\theta}'} .
\end{aligned} \tag{24}$$

The second cross partial derivatives in this expression are

$$\begin{aligned}
\frac{\partial^2 E_1}{\partial r \partial \tilde{\theta}'} &= -\frac{\partial E_1}{\partial r} \left(D_3 + A_0 \frac{\partial a^2(r)}{\partial \tilde{\theta}'} \right) - E_1 \left(\frac{\partial D_3}{\partial r} + \frac{\partial A_0}{\partial r} \frac{\partial a^2(r)}{\partial \tilde{\theta}'} + A_0 \frac{\partial^2 a^2(r)}{\partial r \partial \tilde{\theta}'} \right) \\
&= -A_1 E_1 \left(D_3 + A_0 \frac{\partial a^2(r)}{\partial \tilde{\theta}'} \right) - E_1 \left(A_0 A_1 \frac{\partial a^2(r)}{\partial \tilde{\theta}'} + A_0 \frac{\partial^2 a^2(r)}{\partial r \partial \tilde{\theta}'} \right) \\
\frac{\partial^2 E_3}{\partial r \partial \tilde{\theta}'} &= K_2 \frac{\partial d(r)}{\partial r} F_1 \frac{\partial b(r)}{\partial \tilde{\theta}'} + K_2 d(r) \frac{\partial F_1}{\partial r} \frac{\partial b(r)}{\partial \tilde{\theta}'} + K_2 d(r) F_1 \frac{\partial^2 b(r)}{\partial r \partial \tilde{\theta}'} \\
&\quad + \frac{\partial E_3}{\partial r} \left(Q_1 \frac{\partial Q_1}{\partial \tilde{\theta}'} - D_2 - \frac{1}{2} D_4 \right) \\
&\quad + E_3 \left(\frac{\partial Q_1}{\partial r} \frac{\partial Q_1}{\partial \tilde{\theta}'} + Q_1 \frac{\partial^2 Q_1}{\partial r \partial \tilde{\theta}'} - \frac{\partial D_2}{\partial r} - \frac{1}{2} \frac{\partial D_4}{\partial r} \right)
\end{aligned}$$

$$\begin{aligned}
&= K_2 d(r) Q_1 Q_2 F_1 \frac{\partial b(r)}{\partial \tilde{\theta}'} + \frac{3}{2} K_2 d(r) F_1 A_1 \frac{\partial b(r)}{\partial \tilde{\theta}'} + K_2 d(r) F_1 \frac{\partial^2 b(r)}{\partial r \partial \tilde{\theta}'} \\
&\quad + F_3 \left(Q_1 \frac{\partial Q_1}{\partial \tilde{\theta}'} - D_2 - \frac{1}{2} D_4 \right) + E_3 \left(Q_2 \frac{\partial Q_1}{\partial \tilde{\theta}'} + Q_1 \frac{\partial^2 Q_1}{\partial r \partial \tilde{\theta}'} - \frac{1}{2} \frac{\partial D_4}{\partial r} \right) \\
\frac{\partial^2 P_1}{\partial r \partial \tilde{\theta}'} &= -K_2 Q_1 \exp(-\frac{1}{2} Q_1^2) \frac{\partial Q_1}{\partial r} \frac{\partial Q_1}{\partial \tilde{\theta}'} + K_2 \exp(-\frac{1}{2} Q_1^2) \frac{\partial^2 Q_1}{\partial r \partial \tilde{\theta}'} \\
&= K_2 \exp(-\frac{1}{2} Q_1^2) \left(\frac{\partial^2 Q_1}{\partial r \partial \tilde{\theta}'} - Q_1 Q_2 \frac{\partial Q_1}{\partial \tilde{\theta}'} \right).
\end{aligned}$$

The second cross partial derivatives of $a^2(r)$ and $b(r)$ are

$$\begin{aligned}
\frac{\partial^2 a^2(r)}{\partial r \partial \tilde{\theta}'} &= \left(0, 0, \frac{\rho}{\sigma_1^3 \sigma_2}, \frac{\rho}{\sigma_1 \sigma_2^3} - \frac{2r}{\sigma_2^4}, -\frac{2}{\sigma_1 \sigma_2} \right) \\
B_2 \equiv \frac{\partial^2 b(r)}{\partial r \partial \tilde{\theta}'} &= \left(-\frac{\rho}{\sigma_1 \sigma_2}, \frac{1}{\sigma_2^2}, \frac{\rho \mu_1}{2 \sigma_1^3 \sigma_2}, \frac{\rho \mu_1}{2 \sigma_1 \sigma_2^3} - \frac{\mu_2}{\sigma_2^4}, -\frac{\mu_1}{\sigma_1 \sigma_2} \right).
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\frac{\partial D_4}{\partial r} &= 3A_0 A_1 \frac{\partial a^2(r)}{\partial \tilde{\theta}'} + 3A_0 \frac{\partial^2 a^2(r)}{\partial r \partial \tilde{\theta}'} \\
&= 3A_0 \left(A_1 \frac{\partial a^2(r)}{\partial \tilde{\theta}'} + \frac{\partial^2 a^2(r)}{\partial r \partial \tilde{\theta}'} \right) \\
&= 3A_0 A_2 \\
\frac{\partial^2 Q_1}{\partial r \partial \tilde{\theta}'} &= -\frac{1}{2} \frac{1}{a^3(r) \sqrt{1-\rho^2}} \frac{\partial a^2(r)}{\partial r} \frac{\partial b(r)}{\partial \tilde{\theta}'} + \frac{1}{a(r) \sqrt{1-\rho^2}} \frac{\partial^2 b(r)}{\partial r \partial \tilde{\theta}'} - \frac{1}{2} Q_2 A_0 \frac{\partial a^2(r)}{\partial \tilde{\theta}'} \\
&\quad - \frac{1}{2} Q_1 A_0 A_1 \frac{\partial a^2(r)}{\partial \tilde{\theta}'} - \frac{1}{2} Q_1 A_0 \frac{\partial^2 a^2(r)}{\partial r \partial \tilde{\theta}'} + Q_2 \frac{\rho}{1-\rho^2} e'_5 \\
&= \frac{1}{2} A_1 \frac{1}{a(r) \sqrt{1-\rho^2}} \frac{\partial b(r)}{\partial \tilde{\theta}'} + \frac{1}{a(r) \sqrt{1-\rho^2}} B_2 - \frac{1}{2} Q_2 A_0 \frac{\partial a^2(r)}{\partial \tilde{\theta}'} - \frac{1}{2} Q_1 A_0 A_2 \\
&\quad + Q_2 \frac{\rho}{1-\rho^2} e'_5,
\end{aligned}$$

with A_2 implicitly defined. The expressions for the second cross derivatives of E_1 and E_3 reduce to

$$\frac{\partial^2 E_1}{\partial r \partial \tilde{\theta}'} = -A_1 E_1 \left(D_3 + A_0 \frac{\partial a^2(r)}{\partial \tilde{\theta}'} \right) - E_1 A_0 A_2$$

$$\begin{aligned} \frac{\partial^2 E_3}{\partial r \partial \tilde{\theta}'} &= K_2 d(r) F_1 \frac{\partial b(r)}{\partial \tilde{\theta}'} (Q_1 Q_2 + \frac{3}{2} A_1) + K_2 d(r) F_1 B_2 + F_3 \left(Q_1 \frac{\partial Q_1}{\partial \tilde{\theta}'} - D_2 - \frac{1}{2} D_4 \right) \\ &\quad + E_3 \left(Q_2 \frac{\partial Q_1}{\partial \tilde{\theta}'} + Q_1 \frac{\partial^2 Q_1}{\partial r \partial \tilde{\theta}'} - \frac{3}{2} A_0 A_2 \right), \end{aligned}$$

which can be used in (24).

Derivatives of the mode

From (16), we have that

$$\frac{\partial M}{\partial \theta'} = - \left(\frac{\partial^2 g}{\partial r^2} \right)^{-1} \frac{\partial^2 g}{\partial r \partial \theta'} \Big|_{r=M} = - \left(\frac{\partial^2 g}{\partial r^2} \right)^{-1} \frac{\partial^2 g}{\partial r \partial \tilde{\theta}'} \frac{\partial \tilde{\theta}}{\partial \theta'} \Big|_{r=M}.$$

By inserting (18), (19), (22), and (24) in this expression, we obtain the required derivatives.

Derivatives of the median

The general expression for derivatives of the median is given in (17b). The required derivative of the density function is

$$\frac{\partial g(r)}{\partial \theta'} = \frac{\partial g(r)}{\partial \tilde{\theta}'} \frac{\partial \tilde{\theta}}{\partial \theta'},$$

where the last factor follows from (18) and (19) and the first factor on the right-hand side has been given in (23).

The resulting integral has no closed form solution and there does not seem to be a straightforward way to compute it by numerical integration. Hence, it is most convenient to approximate the integral by simulating a large number of draws from the estimated distribution and replace (17c) by the sample average of the corresponding expression for the simulation sample.

6 Derivatives of the incomplete Beta function

In section 4.3, it was derived that for the derivatives of the median of model IV(c) with respect to the shape parameters α_1 and α_2 , we need the derivatives of the incomplete Beta function ratio

$$I_x(a, b) = \frac{1}{B(a, b)} \int_0^x t^{a-1} (1-t)^{b-1} dt$$

with respect to a and b . We have not been able to find a discussion of the computation of these derivatives in the literature, so we will solve these ourselves. First, note that

$$I_x(a, b) = 1 - I_{1-x}(b, a), \quad (25)$$

an equality that we will use a few times below. Second, it is useful to anticipate the results in section 9: We will need the stated derivatives for values of $x \approx 0.50$ and $a \approx b \approx 1.7$.

Because $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$, it follows that

$$\frac{1}{B(a, b)} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} = \exp[\log \Gamma(a+b) - \log \Gamma(a) - \log \Gamma(b)],$$

and thus

$$\frac{\partial}{\partial a} \frac{1}{B(a, b)} = \frac{1}{B(a, b)} [\Psi(a+b) - \Psi(a)],$$

and analogously $\partial(1/B(a, b))/\partial b = (1/B(a, b))[\Psi(a+b) - \Psi(b)]$, where

$$\Psi(\alpha) = \frac{\partial \log \Gamma(\alpha)}{\partial \alpha} = \frac{1}{\Gamma(\alpha)} \frac{\partial \Gamma(\alpha)}{\partial \alpha}$$

is the digamma function, also called Psi function. Routines for computing the digamma function are widely available, e.g., Bernardo (1976), Amos (1983), or Cody, Strecok, and Thacher (1973), as implemented in Brown et al. (1994). Hence, the problem can be reduced to computing the derivatives of the incomplete Beta function

$$B_x(a, b) = \int_0^x t^{a-1} (1-t)^{b-1} dt$$

with respect to a and b . These are

$$\frac{\partial B_x(a, b)}{\partial a} = \int_0^x t^{a-1} (1-t)^{b-1} \log t \, dt \quad (26a)$$

$$\frac{\partial B_x(a, b)}{\partial b} = - \int_0^x t^{a-1} (1-t)^{b-1} \log(1-t) \, dt. \quad (26b)$$

These integrals do not have a closed-form solution and there does not seem to be a natural way to compute them either. One solution would be to use some form of numerical integration. However, plots of the integrands in (26) show that these change very rapidly over some part of the domain and relatively slowly or steadily over other parts, which may lead to an inaccurate result. By using (25), we may be able to

avoid integrating over the sensitive part, but accurate results will still require heavy computation and therefore, we do not pursue this further.

The integrals in (26) can be expressed as $B(a, b) E[I(t \leq x) \log t]$ and $-B(a, b) E[I(t \leq x) \log(1 - t)]$, respectively, where t follows a Beta distribution with parameters a and b . This suggests that we could simulate a large number of draws from this Beta distribution and use the sample analogs of these expressions for these draws. This will presumably lead to similar numerical inaccuracies (although now stochastic) as direct numerical integration and will therefore not be pursued either.

By expressing $\log(1 - t)$ in a power series expansion in (26b) and interchanging the summation and integration, we obtain a series expansion for the required derivative, but the terms in the expansion are evaluations of the incomplete Beta function themselves, which makes the expression computationally unattractive. Therefore, we do not pursue this further as well.

For the fourth solution, we start with a series expansion of $I_x(a, b)$:

$$I_x(a, b) = \frac{1}{B(a, b)} \frac{x^a}{a} \left[1 + a \sum_{j=1}^{\infty} \frac{\prod_{i=1}^j (i - b)}{j! (a + j)} x^j \right]. \quad (27)$$

A second expression that will be used is

$$\begin{aligned} D_x(a, b, n) &= I_x(a, b) - I_x(a + n, b) \\ &= x^a (1 - x)^b \sum_{j=1}^n \frac{\Gamma(a + b + j - 1)}{\Gamma(b) \Gamma(a + j)} x^{j-1}, \end{aligned} \quad (28)$$

where n is a positive integer. In the region that we are interested in, DiDonato and Morris (1992) use the expressions (27) and (28) to approximate the incomplete Beta function, where the expression used is

$$I_x(a, b) = D_{1-x}(\bar{b}, a, n) + I_x(a, \bar{b}), \quad (29)$$

where $n = \lfloor b \rfloor$, the largest integer smaller than b , and $\bar{b} = b - n$, the remainder. For $D_{1-x}(\bar{b}, a, n)$, an algorithm for computing (28) is used and for $I_x(a, \bar{b})$, the series expansion (27) is truncated at a suitably large value. The derivatives of $D_x(a, b, n)$ with respect to a and b are straightforwardly computed as

$$\begin{aligned} \frac{\partial D_x(a, b, n)}{\partial a} &= D_x(a, b, n) \log x \\ &\quad + x^a (1 - x)^b \sum_{j=1}^n \frac{\Gamma(a + b + j - 1)}{\Gamma(b) \Gamma(a + j)} x^{j-1} \\ &\quad \times [\Psi(a + b + j - 1) - \Psi(a + j)] \end{aligned} \quad (30a)$$

$$\begin{aligned}
\frac{\partial D_x(a, b, n)}{\partial b} &= D_x(a, b, n) \log(1 - x) \\
&+ x^a (1 - x)^b \sum_{j=1}^n \frac{\Gamma(a + b + j - 1)}{\Gamma(b) \Gamma(a + j)} x^{j-1} \\
&\times [\Psi(a + b + j - 1) - \Psi(b)]. \tag{30b}
\end{aligned}$$

The derivative of $I_x(a, b)$ with respect to a in the series expansion (27) is

$$\begin{aligned}
\frac{\partial I_x(a, b)}{\partial a} &= I_x(a, b) [\Psi(a + b) - \Psi(a) + \log x] \\
&- \frac{1}{B(a, b)} \frac{x^a}{a^2} \left[1 + a^2 \sum_{j=1}^{\infty} \frac{\prod_{i=1}^j (i - b)}{j! (a + j)^2} x^j \right]. \tag{31}
\end{aligned}$$

In the region where (27) converges fast, (31) should also converge fast, because the terms in the expansion are smaller due to the square of $(a + j)$ in the denominator. Note that this series is evaluated for $0 \leq b < 1$ and $a > 1$. The error of approximation after truncating the series in (31) after R terms can be bounded as follows. Define

$$c_j = \frac{\prod_{i=1}^j (i - b)}{j! (a + j)^2}.$$

Thus,

$$c_j = c_{j-1} \frac{j - b}{j} \frac{(a + j - 1)^2}{(a + j)^2} < c_{j-1},$$

from which it follows that

$$E_R = \sum_{j=R+1}^{\infty} \frac{\prod_{i=1}^j (i - b)}{j! (a + j)^2} x^j = \sum_{j=R+1}^{\infty} c_j x^j < c_{R+1} \sum_{j=R+1}^{\infty} x^j = c_{R+1} \frac{x^{R+1}}{1 - x}.$$

Consequently, if it is desired that $E_R < \varepsilon$, then we can safely terminate the expansion after R terms if $c_{R+1} x^{R+1} < (1 - x) \varepsilon$.

Analogously, the derivative of $I_x(a, b)$ with respect to b in the series expansion (27) is

$$\frac{\partial I_x(a, b)}{\partial b} = I_x(a, b) [\Psi(a + b) - \Psi(b)] - \frac{1}{B(a, b)} x^a \sum_{j=1}^{\infty} d_j x^j, \tag{32}$$

where

$$d_j = \sum_{i=1}^j \frac{1}{j! (a + j)} \prod_{\substack{k=1 \\ k \neq i}}^j (k - b) = \sum_{i=1}^j \frac{1}{a + j} \frac{1}{i} \prod_{\substack{k=1 \\ k \neq i}}^j \frac{k - b}{k} = \frac{j}{a + j} \frac{1}{j} \sum_{i=1}^j f_{ij}, \tag{33}$$

with f_{ij} implicitly defined. Clearly, if $0 \leq b < 1$, $0 < f_{ij} < 1/i \leq 1$. Moreover, it is easily seen that for these values of b , f_{ij} is decreasing in both i and j and thus $0 < \bar{f}_j \equiv j^{-1} \sum_{i=1}^j f_{ij} < 1$ and \bar{f}_j is decreasing in j . Because $0 < d_j < \bar{f}_j$, (32) always converges for $0 \leq b < 1$. The error of approximation after truncating the series in (32) after R terms can be bounded as follows:

$$\begin{aligned} G_R &= \sum_{j=R+1}^{\infty} d_j x^j < \sum_{j=R+1}^{\infty} \bar{f}_j x^j \\ &< \bar{f}_{R+1} \sum_{j=R+1}^{\infty} x^j = \bar{f}_{R+1} \frac{x^{R+1}}{1-x}. \end{aligned}$$

Consequently, if it is desired that $G_R < \varepsilon$, then we can safely terminate the expansion after R terms if $\bar{f}_{R+1} x^{R+1} < (1-x)\varepsilon$.

By combining (29), (30b), and (31), it follows that the derivative of $I_x(a, b)$ with respect to a can be computed from

$$\begin{aligned} \frac{\partial I_x(a, b)}{\partial a} &= D_{1-x}(\bar{b}, a, n) \log x \\ &+ (1-x)^{\bar{b}} x^a \sum_{j=1}^n \frac{\Gamma(a + \bar{b} + j - 1)}{\Gamma(a) \Gamma(\bar{b} + j)} (1-x)^{j-1} \\ &\quad \times [\Psi(a + \bar{b} + j - 1) - \Psi(a)] \\ &+ I_x(a, \bar{b}) [\Psi(a + \bar{b}) - \Psi(a) + \log x] \\ &- \frac{1}{B(a, \bar{b})} \frac{x^a}{a^2} \left[1 + a^2 \sum_{j=1}^{\infty} \frac{\prod_{i=1}^j (i - \bar{b})}{j! (a + j)^2} x^j \right], \end{aligned} \quad (34)$$

where the series expansion is terminated after a suitable number of terms.

Analogously, by combining (29), (30a), and (32), it follows that the derivative of $I_x(a, b)$ with respect to b can be computed from

$$\begin{aligned} \frac{\partial I_x(a, b)}{\partial b} &= D_{1-x}(\bar{b}, a, n) \log(1-x) \\ &+ (1-x)^{\bar{b}} x^a \sum_{j=1}^n \frac{\Gamma(a + \bar{b} + j - 1)}{\Gamma(a) \Gamma(\bar{b} + j)} (1-x)^{j-1} \\ &\quad \times [\Psi(a + \bar{b} + j - 1) - \Psi(\bar{b} + j)] \\ &+ I_x(a, \bar{b}) [\Psi(a + \bar{b}) - \Psi(\bar{b})] - \frac{1}{B(a, \bar{b})} x^a \sum_{j=1}^{\infty} \bar{d}_j x^j, \end{aligned} \quad (35)$$

with \bar{d}_j as in (33), with \bar{b} instead of b , and where the series expansion is terminated after a suitable number of terms.

Finally, plots of $I_x(a, b)$ as a function of a in the neighborhood of the estimated values of x and b and of $I_x(a, b)$ as a function of b in the neighborhood of the estimated values of x and a show that this function is close to linear for a or b between 0.5 and 2.5, so that numerical derivatives should give good results as well. We use these as a check of the results from (34) and (35).

7 Derivatives for model IV(d)

For model IV(d), the derivatives are quite complicated. It is useful to eliminate the scale parameters from the analysis. As discussed in section 2, we can write $r = (\tau_2/\tau_1)\tilde{r}$ and hence the mean ξ , the variance ω^2 , the mode M , and median m of r can be written in terms of the mean $\tilde{\xi}$, the variance $\tilde{\omega}^2$, the mode \tilde{M} , and the median \tilde{m} of \tilde{r} as

$$\begin{aligned}\xi &= (\tau_2/\tau_1)\tilde{\xi} \\ \omega^2 &= (\tau_2/\tau_1)^2\tilde{\omega}^2 = (\tau_2/\tau_1)^2 \left[\omega_*^2 - \tilde{\xi}^2 \right] \\ M &= (\tau_2/\tau_1)\tilde{M} \\ m &= (\tau_2/\tau_1)\tilde{m},\end{aligned}$$

where $\omega_*^2 = E(\tilde{r}^2)$. The formulas for $\tilde{\xi}$, ω_*^2 , \tilde{M} , and \tilde{m} do not contain τ_1 and τ_2 , which makes the expressions for the derivatives somewhat less complicated. It follows that the derivatives of the mean, standard deviation, mode, and median of r with respect to the parameters $\tilde{\theta}' = (\alpha_1, \alpha_2, \rho)$ are easily obtained from the corresponding derivatives of $\tilde{\xi}$, ω_*^2 , \tilde{M} , and \tilde{m} . As mentioned earlier, the parameter ρ is not estimated directly. Instead, the correlation matrix Σ of the random variables ζ is parameterized as $\Sigma = [\text{diag}(LL')]^{-1/2}LL'[\text{diag}(LL')]^{-1/2}$, where L is a lower triangular matrix with ones on the diagonal. Thus, $\rho = \Sigma_{ij}$ for some $i > j$, which can be further written as

$$\begin{aligned}\rho &= \frac{\sum_{k=1}^j L_{ik}L_{jk}}{\left[\sum_{k=1}^i L_{ik}^2 \sum_{k=1}^j L_{jk}^2 \right]^{1/2}} = \frac{L_{ij} + \sum_{k=1}^{j-1} L_{ik}L_{jk}}{\left\{ \left[1 + \sum_{k=1}^{i-1} L_{ik}^2 \right] \left[1 + \sum_{k=1}^{j-1} L_{jk}^2 \right] \right\}^{1/2}} \\ &= a_{ij}a_{ii}^{-1/2}a_{jj}^{-1/2},\end{aligned}$$

with a_{ij} , a_{ii} , and a_{jj} implicitly defined. It now follows straightforwardly that $\partial\rho/\partial a_{ij} = a_{ii}^{-1/2}a_{jj}^{-1/2}$, $\partial\rho/\partial a_{ii} = -\rho/(2a_{ii})$, $\partial\rho/\partial a_{jj} = -\rho/(2a_{jj})$, $\partial a_{ij}/\partial L_{ij} = 1$, $\partial a_{ij}/\partial L_{ik} = L_{jk}$, and $\partial a_{ij}/\partial L_{jk} = L_{ik}$, where $k < j$. Finally, $\partial a_{ii}/\partial L_{ik} = 2L_{ik}$, where $k < i$, and

$\partial a_{jj}/\partial L_{jk} = 2L_{jk}$, where $k < j$. Using the chain rule now gives the derivatives of ρ with respect to the elements of L .

The derivatives of $\tilde{\xi}$, ω_*^2 , \tilde{M} , and \tilde{m} with respect to $\tilde{\theta}'$ are

$$\begin{aligned}\frac{\partial \tilde{\xi}}{\partial \tilde{\theta}'} &= \int_0^\infty r \frac{\partial \log g(r)}{\partial \tilde{\theta}'} dr \\ \frac{\partial \omega_*^2}{\partial \tilde{\theta}'} &= \int_0^\infty r^2 \frac{\partial \log g(r)}{\partial \tilde{\theta}'} dr \\ \frac{\partial \tilde{M}}{\partial \tilde{\theta}'} &= - \left(\frac{\partial^2 g(r)}{\partial r^2} \right)^{-1} \frac{\partial^2 g(r)}{\partial r \partial \tilde{\theta}'} \Big|_{r=\tilde{M}} \\ &= - \left(\frac{\partial^2 \log g(r)}{\partial r^2} \right)^{-1} \frac{\partial^2 \log g(r)}{\partial r \partial \tilde{\theta}'} \Big|_{r=\tilde{M}} \\ \frac{\partial \tilde{m}}{\partial \tilde{\theta}'} &= - \frac{1}{g(m)} \int_0^m \frac{\partial \log g(r)}{\partial \tilde{\theta}'} dr,\end{aligned}$$

where $g(r)$ is defined in (8). The integrals do not have a closed-form solution. Moreover, because $g(r)$ itself contains an integral that has to be evaluated numerically, it is computationally unattractive to compute these integrals numerically. It is more convenient to compute these “outer” integrals as sample averages of relevant functions based on a simulated sample from the estimated distribution. We will take a somewhat different approach, however.

With a slight abuse of notation, redefine $\theta' = (\alpha_1, \alpha_2, \rho)$ and omit tildes for r , β_1 , and β_2 as well, so that in our new notation,

$$\begin{aligned}r &= \beta_2/\beta_1 \\ \beta_1 &= \beta_1(\zeta_1; \alpha_1) = H^{-1}[\Phi(\zeta_1); \alpha_1] \\ \beta_2 &= \beta_2(\zeta_0, \zeta_1; \alpha_2, \rho) = H^{-1}[\Phi(\zeta_2); \alpha_2] = H^{-1}\left[\Phi\left(\rho\zeta_1 + \sqrt{1-\rho^2}\zeta_0\right); \alpha_2\right],\end{aligned}$$

where ζ_0 and ζ_1 are independent standard normally distributed random variables. Hence, we can write the raw moments of r as

$$\begin{aligned}E_r(r^k) &= E_\zeta[(\beta_2/\beta_1)^k] \\ &= E_\zeta\left[\left(\frac{\beta_2(\zeta_0, \zeta_1; \alpha_2, \rho)}{\beta_1(\zeta_1; \alpha_1)}\right)^k\right] \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left(\frac{\beta_2(\zeta_0, \zeta_1; \alpha_2, \rho)}{\beta_1(\zeta_1; \alpha_1)}\right)^k \phi(\zeta_1)\phi(\zeta_0) d\zeta_0 d\zeta_1,\end{aligned}$$

provided these moments exist. Consequently, the derivatives can be written as

$$\begin{aligned}
\frac{\partial E_r(r^k)}{\partial \theta'} &= E_\zeta \left[k r^{k-1} \frac{\partial r}{\partial \theta'} \right] \\
&= E_\zeta \left[k \left(\frac{\beta_2(\zeta_0, \zeta_1; \alpha_2, \rho)}{\beta_1(\zeta_1; \alpha_1)} \right)^{k-1} \frac{\partial r}{\partial \theta'} \right] \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} k \left(\frac{\beta_2(\zeta_0, \zeta_1; \alpha_2, \rho)}{\beta_1(\zeta_1; \alpha_1)} \right)^{k-1} \frac{\partial r}{\partial \theta'} \phi(\zeta_1) \phi(\zeta_0) d\zeta_0 d\zeta_1.
\end{aligned}$$

The derivatives of r with respect to the elements of θ' are

$$\begin{aligned}
\frac{\partial r}{\partial \alpha_1} &= -\frac{r}{\beta_1} \frac{\partial \beta_1}{\partial \alpha_1} \\
\frac{\partial \beta_1}{\partial \alpha_1} &= -\frac{\partial H(\beta_1; \alpha_1)/\partial \alpha_1}{\partial H(\beta_1; \alpha_1)/\partial \beta_1} = -\Gamma(\alpha_1) \beta_1^{1-\alpha_1} e^{\beta_1} \frac{\partial H(\beta_1; \alpha_1)}{\partial \alpha_1} \\
\frac{\partial r}{\partial \alpha_2} &= \frac{r}{\beta_2} \frac{\partial \beta_2}{\partial \alpha_2} \\
\frac{\partial \beta_2}{\partial \alpha_2} &= -\Gamma(\alpha_2) \beta_2^{1-\alpha_2} e^{\beta_2} \frac{\partial H(\beta_2; \alpha_2)}{\partial \alpha_2} \\
\frac{\partial r}{\partial \rho} &= \frac{r}{\beta_2} \frac{\partial \beta_2}{\partial \rho} \\
\frac{\partial \beta_2}{\partial \rho} &= \frac{\partial \Phi(\zeta_2)/\partial \rho}{\partial H(\beta_2; \alpha_2)/\partial \beta_2} = \Gamma(\alpha_2) \beta_2^{1-\alpha_2} e^{\beta_2} \phi(\zeta_2) \left[\zeta_1 - \frac{\rho}{\sqrt{1-\rho^2}} \zeta_0 \right],
\end{aligned}$$

or, after some simplification,

$$\begin{aligned}
\frac{\partial r}{\partial \alpha_1} &= r \Gamma(\alpha_1) \beta_1^{-\alpha_1} e^{\beta_1} \frac{\partial H(\beta_1; \alpha_1)}{\partial \alpha_1} \\
\frac{\partial r}{\partial \alpha_2} &= -r \Gamma(\alpha_2) \beta_2^{-\alpha_2} e^{\beta_2} \frac{\partial H(\beta_2; \alpha_2)}{\partial \alpha_2} \\
\frac{\partial r}{\partial \rho} &= r \Gamma(\alpha_2) \beta_2^{-\alpha_2} e^{\beta_2} \phi(\zeta_2) \left[\zeta_1 - \frac{\rho}{\sqrt{1-\rho^2}} \zeta_0 \right].
\end{aligned}$$

For the median and the mode, it turns out to be useful to rewrite the distribution function of r in the following form. Let $p = G(q)$ denote a value of the distribution function of r , as a function of a quantile q . Then we can write

$$\begin{aligned}
p &= \Pr(r \leq q) \\
&= \Pr(\beta_2 \leq \beta_1 q) \quad (\text{because } \beta_1 > 0)
\end{aligned}$$

$$\begin{aligned}
&= \Pr [H(\beta_2; \alpha_2) \leq H(\beta_1 q; \alpha_2)] \\
&= \Pr [\Phi(\zeta_2) \leq H(\beta_1 q; \alpha_2)] \\
&= \Pr [\zeta_2 \leq \Phi^{-1}(H(\beta_1 q; \alpha_2))] \\
&= \Pr \left[\rho \zeta_1 + \sqrt{1 - \rho^2} \zeta_0 \leq \Phi^{-1}(H(\beta_1 q; \alpha_2)) \right] \\
&= \Pr \left[\zeta_0 \leq -\frac{\rho}{\sqrt{1 - \rho^2}} \zeta_1 + \frac{1}{\sqrt{1 - \rho^2}} \Phi^{-1}(H(\beta_1 q; \alpha_2)) \right] \\
&= E_{\zeta_1} \left\{ \Phi \left[-\frac{\rho}{\sqrt{1 - \rho^2}} \zeta_1 + \frac{1}{\sqrt{1 - \rho^2}} \Phi^{-1}(H(\beta_1 q; \alpha_2)) \right] \right\}. \tag{36}
\end{aligned}$$

For later use, it is now convenient to define a function $z(x; \alpha) = \Phi^{-1}(H(x; \alpha))$. The derivatives of this function are

$$\begin{aligned}
\frac{\partial z(x; \alpha)}{\partial x} &= \frac{1}{\phi[z(x; \alpha)]} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} \\
\frac{\partial z(x; \alpha)}{\partial \alpha} &= \frac{1}{\phi[z(x; \alpha)]} \frac{\partial H(x; \alpha)}{\partial \alpha}.
\end{aligned}$$

As mentioned above, a routine for computing the derivative $\partial H(x; \alpha)/\partial \alpha$ has been given by Moore (1982). Furthermore, let

$$\zeta_3 = -\frac{\rho}{\sqrt{1 - \rho^2}} \zeta_1 + \frac{1}{\sqrt{1 - \rho^2}} \Phi^{-1}(H(\beta_1 q; \alpha_2)) = \frac{z(\beta_1 q; \alpha_2) - \rho \zeta_1}{\sqrt{1 - \rho^2}} = \frac{\zeta_4 - \rho \zeta_1}{\sqrt{1 - \rho^2}},$$

where $\zeta_4 = z(\beta_1 q; \alpha_2)$. For the derivatives of the median, we now need

$$\frac{\partial q}{\partial \theta'} = -\frac{\partial p / \partial \theta'}{\partial p / \partial q}, \tag{37}$$

evaluated in $q = m$ and $p = 1/2$, cf. (17a). From (36), we have that $p = E_{\zeta_1} [\Phi(\zeta_3)]$, so that

$$\frac{\partial p}{\partial q} = g(q) = E_{\zeta_1}(V_0), \tag{38}$$

where

$$\begin{aligned}
V_0 &= \frac{1}{\sqrt{1 - \rho^2} \Gamma(\alpha_2)} \psi_1 \beta_1^{\alpha_2} q^{\alpha_2-1} e^{-\beta_1 q} \\
\psi_1 &= \phi(\zeta_3)/\phi(\zeta_4) = \exp \left[-\frac{1}{2} \frac{\rho^2 \zeta_1^2 - 2\rho \zeta_1 \zeta_4 + \rho^2 \zeta_4^2}{1 - \rho^2} \right].
\end{aligned}$$

Note that expressions of the form $E_{\zeta_1} [h(\zeta_1)]$ can be easily approximated by simulation of values of ζ_1 from the standard normal distribution, or by rewriting the expression as

$$E_{\zeta_1} [h(\zeta_1)] = \int_{-\infty}^{+\infty} h(\zeta_1) \frac{1}{\sqrt{2\pi}} e^{-\zeta_1^2/2} d\zeta_1 = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} h(x\sqrt{2}) e^{-x^2} dx.$$

and approximating the latter expression by Gauss-Hermite quadrature (see, e.g., Press et al., 1992).

The derivatives in the numerator of (37) are

$$\begin{aligned} \frac{\partial p}{\partial \alpha_1} &= E_{\zeta_1} \left[\phi(\zeta_3) \frac{1}{\sqrt{1-\rho^2}} D_3 B_2 \right] = E_{\zeta_1} [q V_0 B_2], \\ \frac{\partial p}{\partial \alpha_2} &= E_{\zeta_1} \left[\phi(\zeta_3) \frac{1}{\sqrt{1-\rho^2}} B_4 \right] = E_{\zeta_1} \left[\frac{1}{\sqrt{1-\rho^2}} \psi_1 \frac{\partial H(\beta_1 q; \alpha_2)}{\partial \alpha_2} \right], \\ \frac{\partial p}{\partial \rho} &= E_{\zeta_1} \left[\phi(\zeta_3) \frac{1}{(1-\rho^2)^{3/2}} (\rho \zeta_4 - \zeta_1) \right], \end{aligned}$$

where

$$\begin{aligned} D_3 &= \frac{1}{\phi(\zeta_4) \Gamma(\alpha_2)} \beta_1^{\alpha_2} q^{\alpha_2} e^{-\beta_1 q} \\ B_2 &= -\Gamma(\alpha_1) \beta_1^{-\alpha_1} e^{\beta_1} \frac{\partial H(\beta_1; \alpha_1)}{\partial \alpha_1} \\ B_4 &= \frac{1}{\phi(\zeta_4)} \frac{\partial H(\beta_1 q; \alpha_2)}{\partial \alpha_2}. \end{aligned}$$

Combining these expressions with (38) and (37) now gives the derivatives of the median with respect to the parameters.

For the derivatives of the mode, we need the second-order derivatives $\partial^2 g / \partial q^2$ and $\partial^2 g / \partial q \partial \theta'$, which may be derived from (8) or (38). Here, we choose the latter and thus, we can write the derivative of the density as

$$\frac{\partial g}{\partial q} = E_{\zeta_1} \left(\frac{\partial V_0}{\partial q} \right) = E_{\zeta_1} (V_1),$$

where

$$\begin{aligned} V_1 &= \frac{1}{\sqrt{1-\rho^2} \Gamma(\alpha_2)} \psi_1 \beta_1^{\alpha_2} q^{\alpha_2-2} e^{-\beta_1 q} (\alpha_2 - 1 + D_1) \\ D_1 &= \zeta_5 D_3 - \beta_1 q \\ \zeta_5 &= \frac{\rho}{1-\rho^2} (\zeta_1 - \rho \zeta_4). \end{aligned}$$

Similarly,

$$\frac{\partial^2 g}{\partial q^2} = E_{\zeta_1} \left(\frac{\partial V_1}{\partial q} \right) = E_{\zeta_1} (V_2),$$

where

$$\begin{aligned} V_2 &= \frac{1}{\sqrt{1-\rho^2} \Gamma(\alpha_2)} \psi_1 \beta_1^{\alpha_2} q^{\alpha_2-3} e^{-\beta_1 q} [(\alpha_2 - 1 + D_1)(\alpha_2 - 2 + D_1) + D_4] \\ D_4 &= \zeta_7 D_3 - \beta_1 q \\ \zeta_7 &= \zeta_5(\alpha_2 + D_2) - \frac{\rho^2}{1-\rho^2} D_3 \\ D_2 &= \zeta_4 D_3 - \beta_1 q. \end{aligned}$$

Analogously, the second cross-derivative with respect to the quantile q and the parameter α_1 is

$$\frac{\partial^2 g}{\partial q \partial \alpha_1} = E_{\zeta_1} \left(\frac{\partial V_1}{\partial \alpha_1} \right) = E_{\zeta_1} (V_{21}),$$

where

$$V_{21} = \frac{1}{\sqrt{1-\rho^2} \Gamma(\alpha_2)} \psi_1 \beta_1^{\alpha_2} q^{\alpha_2-2} e^{-\beta_1 q} B_2 [(\alpha_2 + D_1)(\alpha_2 - 1 + D_1) + D_4].$$

The second cross-derivative with respect to the quantile q and the parameter α_2 is

$$\frac{\partial^2 g}{\partial q \partial \alpha_2} = E_{\zeta_1} \left(\frac{\partial V_1}{\partial \alpha_2} \right) = E_{\zeta_1} (V_{22}),$$

where

$$\begin{aligned} V_{22} &= \frac{1}{\sqrt{1-\rho^2} \Gamma(\alpha_2)} \psi_1 \beta_1^{\alpha_2} q^{\alpha_2-2} e^{-\beta_1 q} \{ (\alpha_2 - 1 + 2D_1 + \beta_1 q) [\log(\beta_1 q) - \Psi(\alpha_2)] \\ &\quad + 1 + B_4 [\zeta_7 + \zeta_5(D_1 + \beta_1 q - 1)] \}. \end{aligned}$$

Finally, the second cross-derivative with respect to the quantile q and the parameter ρ is

$$\frac{\partial^2 g}{\partial q \partial \rho} = E_{\zeta_1} \left(\frac{\partial V_1}{\partial \rho} \right) = E_{\zeta_1} (V_{23}),$$

where

$$V_{23} = \frac{1}{(1 - \rho^2)^{3/2} \Gamma(\alpha_2)} \psi_1 \beta_1^{\alpha_2} q^{\alpha_2 - 2} e^{-\beta_1 q} [(\alpha_2 - 1 + D_1)(\rho - \psi_3) + \zeta_6 D_3]$$

$$\psi_3 = \frac{1}{1 - \rho^2} (\rho \zeta_4^2 - (1 + \rho^2) \zeta_4 \zeta_1 + \rho \zeta_1^2)$$

$$\zeta_6 = \frac{1}{1 - \rho^2} ((1 + \rho^2) \zeta_1 - 2\rho \zeta_4).$$

Given these formulas, the derivatives of the mode with respect to the parameters now follow straightforwardly from (16).

8 Confidence bands for the density functions and confidence regions for the probability mass function

Occasionally, one may desire to supplement estimated density functions or probability mass functions with some kind of confidence regions as well. Here, it will be argued that these can be obtained using the same principles as discussed above.

First, assume that we have estimated a density function $g(r)$, which implicitly depends on the estimated parameters θ . We can view $g(r)$ for a given value of r as a function of θ , $g(r) = \phi = b(\theta)$, treating r as a known constant. Hence, we can compute $\partial g(r)/\partial \theta$ for each given value of r . Combining this with the delta method, a confidence interval $[g_L(r), g_U(r)]$ for $g(r)$ for a given value of r is easily obtained. We can now view the upper endpoints $g_U(r)$ of the confidence intervals for all r as a function of r again, and similarly for the lower endpoints $g_L(r)$. These functions can then be plotted as (pointwise) confidence bands above and below the estimated density, respectively, thereby giving an impression of the variability of the estimated density functions.

Second, assume that we have estimated a discrete probability mass function, with estimated points of support r_j and probabilities π_j . Conditional on the number J of points of support and the ordering of these points, (r_j, π_j) is a bivariate function of θ . Hence, we can now compute the derivatives $B_j = \partial(r_j, \pi_j)/\partial \theta'$, which can be combined with the delta method to obtain standard errors for r_j and π_j . However, it seems more useful in this case to compute confidence regions for (r_j, π_j) jointly. These can be obtained by inverting a hypothesis test, see, e.g., Wansbeek and Meijer (2000, pp. 294–295). For example, elliptical confidence regions for the probability mass points are obtained as the sets of points (r_j^0, π_j^0) for which Wald statistic

$$T_W = (\hat{r}_j - r_j^0, \hat{\pi}_j - \pi_j^0) \left[\hat{B}_j \hat{V} \hat{B}_j' \right]^{-1} (\hat{r}_j - r_j^0, \hat{\pi}_j - \pi_j^0)'$$

does not exceed the relevant quantile of the χ^2_2 distribution, e.g., 5.99 if the conventional value $\alpha = .05$ is taken.

Although these confidence bands and confidence regions are useful in many practical situations, they are less relevant for our present purposes, because we focus on the differences between different models. Because the functional forms of the densities under the different specifications are different, these density functions are necessarily different and whether or not these differences are within the confidence bands or not tells us more about the sample size than about the differences between the model specifications. With large sample sizes, the functions will become eventually statistically significantly different from each other. For our purposes, it is more relevant to study whether the different model specifications give roughly qualitatively similar density functions, and this subjective judgement is based on inspecting the estimated density functions visually.

9 Empirical results

The formulas for computing the means, standard deviations, medians, modes, and their standard errors have been implemented in an ANSI C program that is available from the author. The program has been extensively tested, especially for parameter values in the neighborhood of the relevant values for our empirical study.

From the tests results, we mention here the testing of the routines for model IV(d). As discussed above, these rely heavily on numerical integration, which can be done by simulation or by Gaussian quadrature. Both options have been tested with $\rho = 0$, so that the results could be compared to the analytical results available for model IV(c). In the simulations, two numbers of replications (drawings) have been tried, namely 100,000 and 1 million. The latter was clearly preferable, but for the precision of the characteristics (mean, standard deviation, median, mode) was still not extremely high, with only 2 or 3 correct significant digits. The quadrature results, with 80 quadrature points, were even worse, however. Of course, more quadrature points would have given better results, but we have not tried this. Consequently, the empirical results presented below have been obtained through simulation with 1 million replications. The precision of the derivatives was generally much better, so standard errors are expected to be suitable.

In table 5, the characteristics and their standard errors are given for the various models, where the computational formulas discussed in this paper have been used. We will not discuss these findings here, a discussion can be found in Meijer and Rouwendal (2004).

Table 5: Characteristics of the distributions of the value of time for the various models.

Model	Mean		Standard deviation		Mode		Median	
Standard logit	0.19	(0.015)	0*	(n.a.)	0.19	(0.015)	0.19	(0.015)
Normal								
(a)	0.26	(0.025)	0.30	(0.017)	0.26	(0.025)	0.26	(0.025)
(b)	0.26	(0.026)	0.31	(0.018)	0.26	(0.026)	0.26	(0.026)
(c)	d.n.e.	(n.a.)	d.n.e.	(n.a.)	0.13	(0.0079)	0.18	(0.011)
(d)	d.n.e.	(n.a.)	d.n.e.	(n.a.)	0.12	(0.0090)	0.18	(0.011)
Lognormal								
(a)	0.37	(0.024)	0.74	(0.11)	0.033	(0.0090)	0.17	(0.017)
(b)	0.48	(0.041)	1.34	(0.30)	0.019	(0.0072)	0.16	(0.018)
(c)	0.35	(0.028)	0.61	(0.087)	0.044	(0.0088)	0.18	(0.015)
(d)	0.35	(0.029)	0.66	(0.10)	0.038	(0.0085)	0.17	(0.015)
Gamma								
(a)	0.28	(0.014)	0.36	(0.018)	d.n.e.	(n.a.)	0.15	(0.018)
(b)	0.36	(0.022)	0.52	(0.031)	d.n.e.	(n.a.)	0.16	(0.024)
(c)	0.45	(0.080)	d.n.e.	(n.a.)	0.046	(0.013)	0.18	(0.015)
(d)	0.56	(0.15)	d.n.e.	(n.a.)	0.045	(0.012)	0.18	(0.014)
Latent class								
(9 classes)	0.42	(0.35)	1.03	(0.94)	0.16	(0.020)	0.16	(0.020)

Note. Standard errors in parentheses.

(a) = independent coefficients, cost coefficient nonrandom,

(b) = dependent coefficients, cost coefficient nonrandom,

(c) = independent coefficients, cost coefficient random as well,

(d) = dependent coefficients, cost coefficient random as well;

* = fixed value; n.a. = not applicable; d.n.e. = does not exist.

10 Existence of characteristics for model IV(d)

Estimation of a characteristic is only sensible when the characteristic actually exists. Of course, the median always exists. But mean, standard deviation, or mode may not exist. For all models except model IV(d) (Gamma model with all coefficients random and dependent), it is known exactly whether these characteristics exist. Sometimes, existence depends on the values of the parameters. See tables 1–3 for details. For model IV(d), however, the precise conditions for existence are not easily derived. Given its similarity with model IV(c), we may expect that the mean and standard deviation exist if α_1 is large enough, and the mode exists if α_2 is large enough. Perhaps the threshold values of α_1 and α_2 for model IV(c) apply as well for model IV(d). Here, we will shed some light on this, but complete “if-and-only-if” results will not be obtained. Note that α_1 and α_2 must be positive to result in proper densities, so this is assumed throughout.

10.1 Existence of expectations

We start with the existence of $E(r^k)$, where k is a positive integer and $r \equiv \tilde{\beta}_2/\tilde{\beta}_1$ as defined in (7). If $k = 1$, this expectation is the mean. The variance and standard deviation of r exist if and only if $E(r^k)$ exists for $k = 2$. So existence of $E(r^k)$, with $k = 1$ or $k = 2$ is of primary interest. For ease of notation, write $x \equiv \tilde{\beta}_1$, $y \equiv \tilde{\beta}_2$, $z \equiv \zeta_1$, $e \equiv \zeta_2 - \rho\zeta_1$, $P_1(t) \equiv H^{-1}[\Phi(t); \alpha_1]$, and $P_2(t) \equiv H^{-1}[\Phi(t); \alpha_2]$. Then z and e are independently normally distributed, $x = P_1(z)$, and $y = P_2(\rho z + e)$. Furthermore, $P_1(t)$ and $P_2(t)$ are monotonically increasing functions of t . Therefore, $P_1^{-1}(u)$ exists and is a monotonically increasing function of u .

We can write

$$E(r^k) = E(y^k x^{-k}) = E_x [x^{-k} E_{y|x}(y^k | x)], \quad (39)$$

provided these expectations exist. Similarly,

$$E(y^k) = \int_0^\infty E_{y|x}(y^k | x), \quad (40)$$

provided these expectations exist. But $E(y^k)$ exists for all positive integers k . Its expression can be given as $E(y^k) = \Gamma(\alpha_2 + k) / \Gamma(\alpha_2) = \alpha_2(\alpha_2 + 1) \cdots (\alpha_2 + k - 1) = \prod_{j=0}^{k-1} (\alpha_2 + j)$. Actually, it exists for all real $k > -\alpha_2$. The first of these expressions can be used for all such k . If $E_{y|x}(y^k | x)$ would not exist for a set (of nonnegative values of x) of positive Lebesgue measure, e.g., an interval of positive length, then the integral in (40) would not exist as well, and thus $E(y^k)$ would not exist. But $E(y^k)$ does exist for the values of k that we are interested in. Consequently, $E_{y|x}(y^k | x)$ exists for all nonnegative x , except possibly for a set of Lebesgue measure zero, i.e., a set of isolated points. Define $Q_k(x) \equiv E_{y|x}(y^k | x)$ if it exists. From

$$Q_k(x) \equiv E_{y|x}(y^k | x) = E_e \left([P_2(\rho P_1^{-1}(x) + e)]^k | x \right), \quad (41)$$

it is evident that $Q_k(x)$ must be continuous and monotonically increasing if $\rho > 0$ and continuous and monotonically decreasing if $\rho < 0$, if it exists. If $\rho = 0$, $Q_k(x) = E(y^k)$ does not depend on x and always exists. Moreover, because y is nonnegative, $Q_k(x)$ is always nonnegative if it exists. Consequently, if $\rho > 0$ and $Q_k(x_0)$ exists for some $x_0 > 0$, then $Q_k(x)$ must exist for all $x \in (0, x_0]$. Conversely, if $Q_k(x_0)$ does not exist for some $x_0 > 0$, then $Q_k(x)$ does not exist for all $x \in [x_0, \infty)$. Similarly, if $\rho < 0$ and $Q_k(x_0)$ exists for some $x_0 > 0$, then $Q_k(x)$ must exist for all $x \in [x_0, \infty)$, and if $Q_k(x_0)$ does not exist for some $x_0 > 0$, then $Q_k(x)$ does not exist for all $x \in (0, x_0]$. It now follows that nonexistence of $Q_k(x_0)$ for some $x_0 > 0$ would imply that the set on

which $Q_k(x)$ would not exist would have positive Lebesgue measure, which is not the case as we have already established above. Hence, $Q_k(x)$ exists for all positive x . The only possible point for which it may not exist is $x = 0$, but this single point will not influence the existence of $E(r^k)$.

We have now argued that $Q_k(x)$ exists for all positive x and is continuous and monotonically increasing if $\rho > 0$, is continuous and monotonically decreasing if $\rho < 0$, and $Q_k(x) = E(y^k)$ does not depend on x if $\rho = 0$. From (39), we have that $E(r^k) = E_x[x^{-k} Q_k(x)]$ if it exists. If $\rho = 0$, $E(r^k) = E(y^k) E(x^{-k})$ exists if and only if $E(x^{-k})$ exists, because $E(y^k)$ always exists for the values of k that we are interested in. It is well known (and can be easily derived from the Gamma integral) that $E(x^{-k})$ exists if and only if $\alpha_1 - k > 0$, i.e., $\alpha_1 > k$. The model with $\rho = 0$ is model IV(c), so this confirms the existence conditions for model IV(c) given in tables 1 and 2.

Positive dependence

Let $f_x(x)$ be the (marginal) density function of x . When $\rho > 0$, $Q_k(x)$ is a monotonically increasing function of x . Consequently, if $x \in (0, 1)$ then $Q_k(x) < Q_k(1)$ and thus $x^{-k} Q_k(x) < x^{-k} Q_k(1)$. Hence, $\int_0^1 x^{-k} Q_k(x) f_x(x) dx$ exists if $\int_0^1 x^{-k} f_x(x) dx$ exists. And this obviously exists if $E(x^{-k}) = \int_0^\infty x^{-k} f_x(x) dx$ exists. If $x \in (1, \infty)$ then $x^{-k} < 1$ and thus $x^{-k} Q_k(x) < Q_k(x)$. Hence, $\int_1^\infty x^{-k} Q_k(x) f_x(x) dx$ exists if $\int_1^\infty Q_k(x) f_x(x) dx$ exists. And this obviously exists if $E(y^k) = \int_0^\infty Q_k(x) f_x(x) dx$ exists. But the existence of the latter has already been mentioned above. Therefore, if $E(x^{-k})$ exists, i.e., if $\alpha_1 > k$, then $E(r^k)$ also exists. The existence of $E(x^{-k})$ is a sufficient condition for the existence of $E(r^k)$. It may, however, not be necessary. If there exists a (small) positive number δ such that $x^{-k} Q_k(x) < Mx^{-\alpha_1}$ for all $x \in (0, \delta)$, i.e., if

$$Q_k(x) < Mx^{k-\alpha_1} \quad (42)$$

for all $x \in (0, \delta)$, where M is a finite constant not depending on x , but possibly depending on δ , then $E(r^k)$ still exists. With $\alpha_1 < k$, both sides of the inequality in (42) are increasing functions of x , so the condition could be satisfied for certain values of $(\alpha_1, \alpha_2, \rho)$, but we have not been able to prove or disprove this. The existence can, however, be studied empirically for given values of $(\alpha_1, \alpha_2, \rho)$, see below.

Negative dependence

When $\rho < 0$, we have the mirror image of the previous case. Let us first study the problem area for the existence of $E(x^{-k})$. If this expectation exists, it can be written as

$$\begin{aligned} E(x^{-k}) &= \int_0^\infty \frac{1}{\Gamma(\alpha_1)} x^{\alpha_1-k-1} e^{-x} dx \\ &= \int_0^1 \frac{1}{\Gamma(\alpha_1)} x^{\alpha_1-k-1} e^{-x} dx + \int_1^\infty \frac{1}{\Gamma(\alpha_1)} x^{\alpha_1-k-1} e^{-x} dx. \end{aligned}$$

If $x > 1$, $x^{-k} < 1$, so

$$\int_1^\infty \frac{1}{\Gamma(\alpha_1)} x^{\alpha_1-k-1} e^{-x} dx < \int_1^\infty \frac{1}{\Gamma(\alpha_1)} x^{\alpha_1-1} e^{-x} dx < \int_0^\infty \frac{1}{\Gamma(\alpha_1)} x^{\alpha_1-1} e^{-x} dx = 1,$$

so the integral $\int_1^\infty \frac{1}{\Gamma(\alpha_1)} x^{\alpha_1-k-1} e^{-x} dx$ always exists. Consequently, if $E(x^{-k})$ does not exist, this is due to nonexistence of the integral $\int_0^1 \frac{1}{\Gamma(\alpha_1)} x^{\alpha_1-k-1} e^{-x} dx$.

With $\rho < 0$, $Q_k(x)$ is a monotonically decreasing function of x . Consequently, if $x \in (1, \infty)$ then $Q_k(x) < Q_k(1)$ and thus $\int_1^\infty x^{-k} Q_k(x) f_x(x) dx < Q_k(1) \int_1^\infty x^{-k} f_x(x) dx$, which always exists, as argued above. So this part of the integral does not influence existence or nonexistence of $E(r^k)$. If $x \in (0, 1)$ then $Q_k(x) > Q_k(1)$ and thus $x^{-k} Q_k(x) > x^{-k} Q_k(1)$. Hence, $\int_0^1 x^{-k} Q_k(x) f_x(x) dx$ does not exist if $\int_0^1 x^{-k} f_x(x) dx$ does not exist. From the analysis above, it follows that this is equivalent to nonexistence of $E(x^{-k})$. Therefore, if $E(x^{-k})$ does not exist, i.e., if $\alpha_1 < k$, then $E(r^k)$ does not exist as well. Therefore, existence of $E(x^{-k})$ is a necessary condition for the existence of $E(r^k)$. It may, however, not be sufficient. If there does not exist a (small) positive number δ such that $x^{-k} Q_k(x) < Mx^{-\alpha_1}$ for all $x \in (0, \delta)$, where M is a finite constant not depending on x , but possibly depending on δ , i.e., if

$$\lim_{x \downarrow 0} x^{\alpha_1-k} Q_k(x) = +\infty \quad (43)$$

then $E(r^k)$ does not exist. With $\alpha_1 > k$, the first factor on the left-hand side of (43) is an increasing function of x and the second factor is a decreasing function of x , so the condition could be satisfied for certain values of $(\alpha_1, \alpha_2, \rho)$ and not for others, but we have not been able to prove or disprove this.

However, the behavior of the left-hand side of (43) near zero can be studied empirically for given parameter values. For a given value of x , $Q_k(x)$ can be approximated by numerical integration, such as quadrature or simulation. In the latter case, computations are based on (41), which is an expectation of a complicated, but computable, function of a normal random variable e with mean zero and variance

$(1 - \rho^2)$, and the given value of x . In this way, $x^{\alpha_1-k} Q_k(x)$ can be computed for small values of x , from which it can be tentatively concluded whether the limit is finite and thus $E(r^k)$ exists, or the limit is infinite and thus $E(r^k)$ does not exist. Moreover, the condition (43) applies to $\rho > 0$ as well, so this procedure can also be used for all values of ρ .

We have applied this procedure for the empirical example discussed in the text. For this example, $\alpha_1 = 1.36$, $\alpha_2 = 1.67$, and $\rho = 0.085$. Based on the analysis above, the mean exists because $\alpha_1 > 1$ and $\rho > 0$. These are sufficient conditions. With $\rho > 0$, a sufficient condition for the existence of the variance would be $\alpha_1 > 2$, but this is not satisfied. If we take $\alpha_1 > 2$ as a simple rule of thumb for existence, we suspect that the variance does not exist.

In figures 1 and 2, the functions $R_k(x) \equiv x^{\alpha_1-k} Q_k(x)$, with $k = 1$ and $k = 2$, are plotted against x for small values of x . Note that both figures are plotted on a double logarithmic scale. Clearly, figure 1 confirms the existence of the mean and figure 2 confirms our suspected nonexistence of the variance. Moreover, these figures strongly suggest that $\log R_k(x)$ is linear in $\log x$ for small values of x , which implies that $\log Q_k(x) = A_k + B_k \log x$ for some constants A_k and B_k , which could be estimated by linear regression. We have not done this, nor do we need this. However, if it could be proven that this formula holds for all values of the parameters (where the constants A_k and B_k will depend on the parameters), then proving that $E(r^k)$ exists or does not exist will be easy.

10.2 Existence of the mode

The typical way to check whether the mode of r exists is to compute the first derivative of the density function with respect to r and study whether it is zero for some value of r , and if so, for which value. Unfortunately, for model IV(d), the first derivative of the density is a complicated function that contains integrals that have no closed-form solution. These expressions do not give much insight in whether the mode exists. Therefore, we will take a different approach here.

The density function $g(r)$ is continuous for all $r \in (0, \infty)$, and it can only be a proper density if $g(r) \rightarrow 0$ when $r \rightarrow \infty$. If $\lim_{r \downarrow 0} g(r)$ is finite, redefine $g(0)$ to be this limit if $g(0)$ does not exist. Then the continuity of $g(r)$ assures that this density exists and is finite for all nonnegative r . The result that $g(r) \rightarrow 0$ when $r \rightarrow \infty$ implies that there exists a (possibly large) number $M > 0$ such that $g(r) < g(M)$ for all $r > M$. Therefore, the area $r > M$ does not influence existence or nonexistence of the mode and can thus be safely ignored. Because the density $g(r)$ is finite and continuous on the closed interval $[0, M]$, it must attain its maximum somewhere in this interval and thus the mode exists when $\lim_{r \downarrow 0} g(r)$ is finite. Conversely, obviously the mode does not

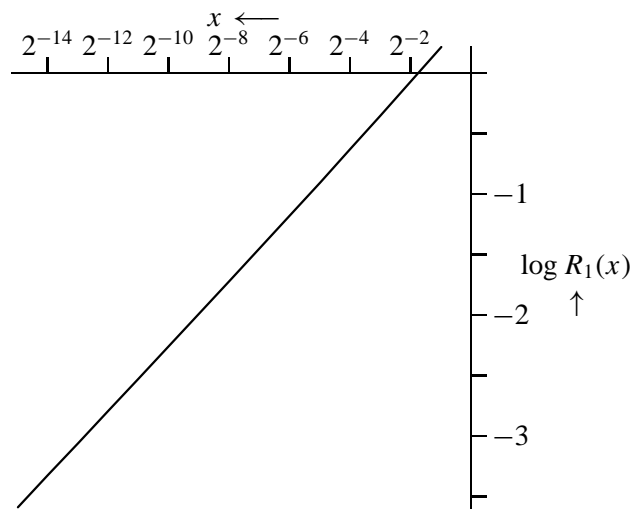


Figure 1: The function $R_1(x)$ for the empirical example.

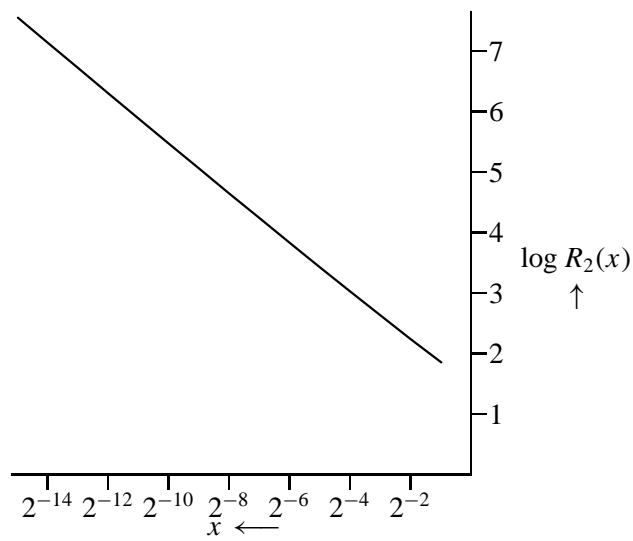


Figure 2: The function $R_2(x)$ for the empirical example.

exist if $\lim_{r \downarrow 0} g(r)$ is infinite. Apparently, the behavior of the density when r is close to zero is conclusive about the existence of the mode.

The density function has been given in (8), which we repeat here:

$$g(r) = C(\alpha_1, \alpha_2, \rho) \frac{r^{\alpha_2-1}}{(1+r)^{\alpha_1+\alpha_2}} \int_0^{+\infty} v^{\alpha_1+\alpha_2-1} e^{-v} h(v; r; \alpha_1, \alpha_2, \rho) dv.$$

The limit as $r \downarrow 0$ is determined by the factors r^{α_2-1} and $h(v; r; \alpha_1, \alpha_2, \rho)$. Clearly, if the latter is bounded away from zero and infinity for small r , the former determines the required limit, and it follows that the mode exists if and only if $\alpha_2 \geq 1$. Unfortunately, however, the behavior of $h(v; r; \alpha_1, \alpha_2, \rho)$ is difficult to assess. Moreover, it may be incorrect to take its limit for $r \downarrow 0$ (if this could be computed), because it must be integrated over v first, and it is far from guaranteed that the limit of the integral and the integral of the limit produce the same results.

On the other hand, the density function can be computed numerically, as discussed above in sections 2 and 7. Then, we can study the behavior of the density near zero empirically in the same way we studied the existence of the mean and variance above. Figure 3 shows the density for small values of r , again on a double logarithmic scale. It clearly shows that the density decreases towards zero with $r \downarrow 0$. For small values of r (say $r < 2^{-6}$), the figure strongly suggests that $\log g(r)$ is linear in $\log r$. Furthermore, the figure shows that the mode must be between $2^{-3} = 0.125$ and $2^{-1} = 0.5$ (the density was only computed for negative integer powers of 2 for the figure). The more precise computations in section 9 give the result $\text{mode} = 0.045$ for the scaled variable, i.e., multiplied by $(\tau_2/\tau_1) = (-0.09788948/-0.7183382) = 0.13627$. Retransformation to the current (standardized) scale, the mode in the figure is 0.33. The density function of the unstandardized r is depicted on a linear scale for a much larger range of r values in Meijer and Rouwendal (2000, 2004).

References

- Amos, D. E. (1983). Algorithm 610: A portable FORTRAN subroutine for derivatives of the psi function. *ACM Transactions on Mathematical Software*, 9, 494–502.
- Aptech Systems. (1998). GAUSS for Windows NT/95 Version 3.2 [Computer software]. Maple Valley, WA: Aptech Systems.
- Bernardo, J. M. (1976). Algorithm AS 103: Psi (digamma) function. *Applied Statistics*, 25, 315–317.
- Berndt, E. K., Hall, B. H., Hall, R. E., & Hausman, J. A. (1974). Estimation and inference in non-linear structural models. *Annals of Economic and Social Measurement*, 3, 653–665.

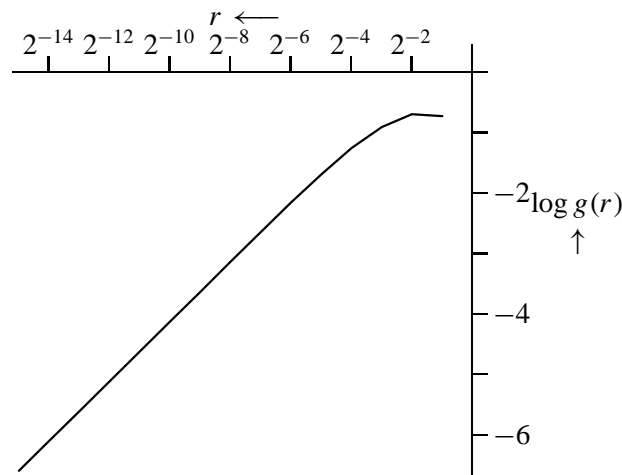


Figure 3: The density function $g(r)$ for the empirical example.

- Brown, B. W., Lovato, J., & Russell, K. (1994). Dcdfliib [collection of C routines]. Available from <http://www.netlib.org>
- Cody, W. J., Strecok, A. J., & Thacher, H. C., Jr. (1973). Chebyshev approximations for the psi function. *Mathematics of Computation*, 27, 123–127.
- Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge, UK: Cambridge University Press.
- Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society B*, 52, 393–442. (with discussion)
- DiDonato, A. R., & Morris, A. H., Jr. (1986). Computation of the incomplete gamma function ratios and their inverse. *ACM Transactions on Mathematical Software*, 12, 377–393.
- DiDonato, A. R., & Morris, A. H., Jr. (1987). Algorithm 654: FORTRAN subroutines for computing the incomplete gamma function ratios and their inverse. *ACM Transactions on Mathematical Software*, 13, 318–319.
- DiDonato, A. R., & Morris, A. H., Jr. (1992). Algorithm 708: Significant digit computation of the incomplete beta function ratios. *ACM Transactions on Mathematical Software*, 18, 360–373.
- Hogg, R. V., & Klugman, S. A. (1983). On the estimation of long tailed skewed distributions with actuarial applications. *Journal of Econometrics*, 23, 91–102.
- Johnson, N. L., & Kotz, S. (1970). *Distributions in statistics: Continuous univariate*

- distributions-2*. New York: Wiley.
- Marsaglia, G. (1965). Ratios of normal variables and ratios of sums of uniform variables. *Journal of the American Statistical Association*, 60, 193–204.
- MathWorks. (1999). Matlab, Version 5.3 [Computer software]. Natick, MA: MathWorks.
- Meijer, E. (1998). *Structural equation models for nonnormal data*. Leiden: DSWO Press.
- Meijer, E., & Rouwendal, J. (2000). *Measuring welfare effects in models with random coefficients* (Research Report No. 00F25). Groningen: SOM Research School, University of Groningen.
- Meijer, E., & Rouwendal, J. (2004). *Measuring welfare effects in models with random coefficients*. Manuscript submitted for publication.
- Moore, R. J. (1982). Algorithm AS 187: Derivatives of the incomplete gamma integral. *Applied Statistics*, 31, 330–335.
- Moran, P. A. P. (1969). Statistical inference with bivariate gamma distributions. *Biometrika*, 56, 627–634.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in FORTRAN: The art of scientific computing* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Rao, C. R. (1973). *Linear statistical inference and its applications* (2nd ed.). New York: Wiley.
- Scales, L. E. (1985). *Introduction to non-linear optimization*. London: Macmillan.
- Shea, B. L. (1988). Algorithm AS 239: Chi-squared and incomplete gamma integral. *Applied Statistics*, 37, 466–473.
- SPSS. (1998). SPSS for Windows 9.0 [Computer software]. Chicago: SPSS.
- Wansbeek, T., & Meijer, E. (2000). *Measurement error and latent variables in econometrics*. Amsterdam: North-Holland.